

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 13.078 Seconds
(without alignments)
1283.801 Million cell updates/sec

Title: US-10-007-267-2
Perfect score: 598
Sequence: 1 LQAVAVFKQLPEAAALAAAN.....LLNRLAEQMNVAADIALIGE 126

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	598	100.0	126	18	AAW06577
2	164	27.4	672	23	ABBS4412
3	163	27.3	989	22	AAU33401
4	161	26.9	678	18	AAW33905
5	161	26.9	678	19	AAW37953
6	154	25.8	679	23	ABP26830
7	135	22.6	100	23	ABP26829
8	135	22.6	679	23	ABP28806
9	132.5	22.2	80	23	ABP07415
10	119	19.9	1067	23	ABBS2588

11	103.5	17.3	688	23	ABBA9111
12	88	14.7	1503	20	AAV27142
13	88	14.7	1503	20	AAW6701
14	88	14.7	1525	20	AAV17499
15	88	14.7	1529	20	AAV27145
16	88	14.7	1529	20	AAW6702
17	88	14.7	1529	22	AAW61239
18	87	14.5	1529	21	AAV76117
19	87	14.5	1529	22	AAW5056
20	87	14.5	1529	23	ABW7236
21	85	14.2	1530	21	AAW28151
22	83	13.9	599	22	ABG24227
23	82.5	13.8	484	22	AAW78985
24	82.5	13.8	533	22	AAW79699
25	82.5	13.8	620	21	AAW51261
26	82.5	13.8	638	21	AAW51260
27	82.5	13.8	754	21	AAW51259
28	82.5	13.8	1374	22	AAW69070
29	82	13.7	232	21	AAV74904
30	82	13.7	232	21	AAV74907
31	82	13.7	815	23	ABP27680
32	82	13.7	815	23	ABP29822
33	80.5	13.5	232	21	AAV74906
34	80.5	13.5	1224	22	ABW62191
35	80	13.4	231	21	AAV74903
36	79.5	13.3	594	21	AAV44280
37	79.5	13.3	647	21	AAV44279
38	79.5	13.3	647	21	AAV44285
39	79.5	13.3	647	21	AAV44286
40	79.5	13.3	647	21	AAV44287
41	79.5	13.3	647	21	AAV44288
42	79	13.2	888	23	AAW83013
43	78.5	13.1	255	18	AAW20772
44	78.5	13.1	1030	21	AAW38915
45	78.5	13.1	1133	21	AAW38914

ALIGNMENTS

RESULT 1	
AAW06577	AAW06577 standard; Protein: 126 AA.
XX	XX
XX	AAW06577;
XX	21-MAR-1997 (first entry)
DT	DT
XX	XX
XX	Lipo-oligosaccharide gene-encoded protein.
DE	DE
XX	XX
XX	Polyglycosyltransferase; N-acetylglucosaminyl transferase;
KW	KW N-acetylglucosaminyl transferase; lipo-oligosaccharide.
XX	XX
OS	OS Neisseria gonorrhoeae ATCC 33084.
XX	XX
PN	PN W09640971-A1.
XX	XX
PD	PD 19-DEC-1996.
XX	XX
PF	PF 03-JUN-1996; 96W0-US08323.
XX	XX
PR	PR 07-JUN-1995; 95US-0478140.
XX	XX
PA	PA (NEOS-) NEOSE TECHNOLOGIES INC.
XX	XX
PI	PI Buczala SL, Johnson KF, Roth S;
XX	XX
DR	DR WPI: 1997-052351/05.
XX	XX
DR	DR N-PSDB: AAT49230.
XX	XX
PT	PT Transfer of at least 2 saccharide units using
PT	PT poly:glycosyltransferase - isolated from N. gonorrhoeae, catalyses
PT	PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a

Listeria monocytog
Human silt-2 matur
Silt-like protein
Human Silt-1 prote
Human silt-2 prote
Full length silt-1
Human Silt-1 prote
Rat Silt homologue
Skin cell protein,
Rat protein isolat
Silt protein. Xen
Novel human diagno
Human protein SEO
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human male enhance
Neisseria meningit
Neisseria meningit
Streptococcus poly
Streptococcus poly
Neisseria meningit
Drosophila melanog
Neisseria meningit
Murine A20 Binding
Murine A20 Binding
Mutant A20 Binding
Mutant A20 Binding
Mutant A20 Binding
Human A20 Binding
Human homologue of
H. pylori cytoplas
Arabidopsis thalia

PT single galactose moiety
 XX
 PS Disclosure; Fig 2A; 38pp; English.
 XX
 CC A lipooligosaccharide-encoding gene region (AAT49230) of *Neisseria*
 CC gonorrhoeae ATCC 33084 includes coding sequences for 5 proteins
 CC (AAW0576-80), one of which (AAW0576) is a polyglycosyltransferase
 CC that catalyses the addition of GlcNAc and GalNAc disaccharides to
 CC a galactose moiety. The function of the other 4 proteins is not
 CC stated in the specification.
 XX
 SQ Sequence 126 AA;
 Query Match 100.0%; Score 598; DB 18; Length 126;
 Best Local Similarity 100.0%; Pred. No. 1.3e-55;
 Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LQAAVAFKQLEPAAALAAANKRVONLKKADALGEVNESLLOODEERKALYAAAGLOPK 60
 DB 1 LQAAVAFKQLEPAAALAAANKRVONLKKADALGEVNESLLOODEERKALYAAAGLOPK 60
 QY 61 IAAVABEGNFTALSELASVSRPOVDAPFDGYVMVAEDAAVAKONRLNLRLAEQNNAYAD 120
 DB 61 IAAVABEGNFTALSELASVSRPOVDAPFDGYVMVAEDAAVAKONRLNLRLAEQNNAYAD 120
 QY 121 IALLGE 126
 DB 121 IALLGE 126
 Db 121 IALLGE 126
 RESULT 2
 ABB54412
 ID ABB54412 standard; Protein: 672 AA.
 XX
 AC ABB54412;
 DT 16-MAY-2002 (first entry)
 XX
 DE Lactococcus lactis protein glusB.
 XX
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX
 OS Lactococcus lactis IL1403.
 XX
 PS FR2807446-A1.
 PD 12-OCT-2001.
 XX
 PF 11-APR-2000; 2000FR-0004630.
 XX
 PR 11-APR-2000; 2000FR-0004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 DR WPI; 2002-043418/06.
 XX
 PT New nucleotide sequence useful in the identification or Lactococcus
 PT lactis and related species -
 XX
 PS Claim 6; SEQ ID No 1114; 2504pp; French.
 XX
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABB53300-ABB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO20017734 (published 18-OCT-2001) which is available in electronic

CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 672 AA;
 Query Match 27.4%; Score 164; DB 23; Length 672;
 Best Local Similarity 38.0%; Pred. No. 9.6e-09;
 Matches 38; Conservative 21; Mismatches 37; Indels 4; Gaps 2;
 QY 25 NLKKADALGEVNESLLOODEERKALYAAAGLOPKIAAAVABEGNFTALSELASVSRPOV 84
 DB 575 NLVAKA-KDIKEINPTLFEEDAEALYNNVISLONQTYMPGEKFAIYHSLA---PAI 630
 QY 85 DAFPDGYVMVAEDAAVAKONRLNLRLAEQNNAYADIAL 124
 DB 631 ETFESVYMAEDLSVRDNRITALLSEVVALTSVMADESLI 670
 RESULT 3
 AU33401
 ID AU33401 standard; Protein: 989 AA.
 XX
 AC AU33401;
 DT 14-FEB-2002 (first entry)
 XX
 DE Enterococcus faecalis cellular proliferation protein #37.
 XX
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO200170955-A2.
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 XX
 PR 23-MAY-2000; 2000US-206848P.
 XX
 PR 26-MAY-2000; 2000US-207727P.
 XX
 PR 23-OCT-2000; 2000US-242578P.
 XX
 PR 27-NOV-2000; 2000US-253625P.
 XX
 PR 22-DEC-2000; 2000US-257931P.
 XX
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 XX
 DR N-PSDB; AAS51260.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 4897; 51pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an

CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 989 AA:

Query Match 27.3%; Score 163; DB 22; Length 989;
Best Local Similarity 31.5%; Pred. No. 2e-08; Indels 26; Gaps 3;
Matches 46; Conservative 19; Mismatches 55;

QY 1 LGAVAVFQKLPFAALAAAN-----KRVNLTKKADALGEVNE--- 39
DB 848 IDAVVSAEQADSLKFAASANIILKSRFDDQDFKPSMEALTRVNLAKKGGELLDGDFEGID 907
QY 40 -SLQODEKALYAAAGCIGPKIAAAGNFRRTALSELASVKPOVDAFPDGVMAEDA 98
DB 908 PSLEFKAKKELDYQVNDLSEAFRTIENYEAE---LVNRLPLIDAYFNETMVAEDE 963
QY 99 AVKONRLNLNRLAEQMAVADIAL 124
DB 964 KYKONRLKOLMOIKAMALSTASLDL 989

RESULT 4

AAM33905

ID AAM33905 standard: Protein; 678 AA.

AC AAM33905;

DT 28-APR-1998 (first entry)

XX Streptococcus pneumoniae glycy1 tRNA synthetase beta.

XX Glycy1 tRNA synthetase beta; glyys beta polypeptide; immune response;

KW vaccine; antibacterial; meningitis; bacterial infection.

OS Streptococcus pneumoniae.

XX MO9738716-A1.

XX 23-OCT-1997.

XX 18-APR-1997; 97WO-US06552.

XX 18-APR-1996; 96GB-0008002.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX Lawlor EJ;

XX WPI; 1997-526209/48.

XX N-PSDB; AAT92936.

PT DNA encoding glycy1 tRNA synthetase of Streptococcus pneumoniae -
PT useful for diagnosis, treatment and prevention of bacterial
PT infections, especially meningitis

PS Claim 12; Page 40-42; 51pp; English.

XX This sequence represents the glycy1 tRNA synthetase (glyys) beta subunit
CC from Streptococcus pneumoniae 0100993 (NCIMB 40794). The polynucleotide
CC encoding glyys can be used for the recombinant production of glyys, which
CC can be used to treat conditions requiring glyys activity. Sequences
CC antisense to the polynucleotide can be used to control glyys expression.
CC glyys, or glyys expressing vectors can be used to induce an immune
CC response, i.e. an antibody (Ab) and/or T cell response, against
CC S. pneumoniae to protect against infection, or to screen for antagonists
CC or agonists of the polynucleotide encoding glyys or glyys's activity,
CC i.e. antibacterials. The antagonist, e.g. an anti-glyys Ab, can be used to
CC treat conditions requiring glyys inhibition, e.g. an S. pneumoniae

CC infection, particularly meningitis. A fragment of the polynucleotide
CC encoding glyys can be used as a probe to isolate full length or related
CC sequences, or diagnose, e.g. by polymerase chain reaction, infection
CC stage and type, including mutation and polymorphism detection. Diagnosis
CC may also be achieved by detecting glyys gene overexpression, e.g. by
CC immunosassay. The Ab can be used to treat infection, isolate or identify glyys
CC expressing clones, purify glyys and as an immunosassay reagent. More
CC generally, the products can prevent adhesion of bacteria to wounds and in
CC dwelling devices, block glyys protein mediated invasion of mammalian cells
CC and block the normal progression of infection.

SQ Sequence 678 AA:

Query Match 26.9%; Score 161; DB 18; Length 678;
Best Local Similarity 35.5%; Pred. No. 2e-08; Indels 20; Gaps 3;
Matches 43; Conservative 21; Mismatches 37;

QY 6 VFQKLPFAAALAAANK-----RVQNLTKKADALGEVNESLQODEKALYAA 53
DB 553 VADMLEAASALVEYSKEEDFKPSVESLSRAFNIAEKAE-GVATVDSALFENDQEKALAA 611
QY 54 AAGLOPKIAAAGNFRRTALSELASVKPOVDAFPDGVMAEDAIVKONRLNLNRLAE 113
DB 612 VETL-----VLSGPAASQOLKOLFALSPVIDAFENTVMAEDQAVRONRLAISQLTK 664
QY 114 Q 114
DB 665 K 665

RESULT 5

AAM37953

ID AAM37953 standard: Protein; 678 AA.

AC AAM37953;

DT 21-AUG-1998 (first entry)

XX Amino acid sequence of glyys (beta) protein.

XX Glycy1 tRNA synthetase-beta gene; glyys-beta; vaccine; agonist;

KW antagonist; antibacterial activity.

OS Streptococcus pneumoniae.

XX US5756330-A.

XX 26-MAY-1998.

XX 18-APR-1997; 97US-0844085.

XX 18-APR-1997; 97US-0844085.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX Lawlor EJ;

XX WPI; 1998-321534/28.

XX N-PSDB; AAV31656.

PT DNA encoding pneumococcal glycy1 tRNA synthetase subunit
PT polypeptides - useful for producing recombinant polypeptides,
PT screening for antibacterial compounds, and in DNA vaccines

PS Claim 1; Columns 7-8; 24pp; English.

XX This is the amino acid sequence of the Streptococcus pneumoniae novel
CC glycy1 tRNA synthetase-beta (glyys-beta) protein. The polynucleotides
CC are useful for producing the recombinant polypeptides, which can be
CC used in vaccines or to screen for glyys-beta or glyys-alpha agonists or
CC antagonists with potential antibacterial activity. The
CC polynucleotides are also useful for DNA vaccination.

CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 XX Sequence 100 AA;
 SQ
 Query Match 22.6%; Score 135; DB 23; Length 100;
 Best Local Similarity 35.0%; Pred. No. 9.9e-07;
 Matches 36; Conservative 18; Mismatches 41; Indels 8; Gaps 2;
 OY 22 RVQNLKKADALGEVNESLQODEKALYAAAGLOPKIAAAGEGFTALSELASVK 81
 DB 3 RAFNLAEKVTHSV-LVDSLSFENNQEKALYQAILSL-----LTEDMHNLDKIFALS 54
 OY 82 PQVDAFEDGVVMAEDAANKONRLNLRLAEQNNVADIALL 124
 DB 55 PIIDFEDNTWMTDDEKMKONRLAILNSIVAKARTVAAAFNML 97
 RESULT 8
 ABP28806
 ID ABP28806 standard; Protein: 679 AA.
 XX
 AC ABP28806;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 6788.
 XX
 KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KM group A streptococcus; Streptococcus pyogenes; antibacterial;
 KM antiinflammatory; Infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus agalactiae.
 XX
 PN W0200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tetelin H;
 XX
 DR WPI; 2002-352536/38.
 DR N-PSDB; ABN69437.
 XX
 XX New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 Claim 1; Page 3841; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 XX Sequence 679 AA;
 SQ
 Query Match 22.6%; Score 135; DB 23; Length 679;
 Best Local Similarity 35.0%; Pred. No. 1.1e-05;
 Matches 36; Conservative 18; Mismatches 41; Indels 8; Gaps 2;
 OY 22 RVQNLKKADALGEVNESLQODEKALYAAAGLOPKIAAAGEGFTALSELASVK 81
 DB 582 RAFNLAEKVTHSV-LVDSLSFENNQEKALYQAILSL-----LTEDMHNLDKIFALS 633
 OY 82 PQVDAFEDGVVMAEDAANKONRLNLRLAEQNNVADIALL 124
 DB 634 PIIDFEDNTWMTDDEKMKONRLAILNSIVAKARTVAAAFNML 676
 RESULT 9
 ABP07415
 ID ABP07415 standard; Protein: 80 AA.
 XX
 AC ABP07415;
 XX
 DT 24-JUN-2002 (first entry)
 XX
 DE Human ORFX protein sequence SEQ ID NO:14812.
 XX
 KM Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KM hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KM degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KM hypertension; hypothyroidism; cholesterol ester storage disease;
 KM immune deficiency; immune disorder; infectious disease;
 KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KM myasthenia gravis.
 XX
 OS Homo sapiens.
 XX
 PN W0200192523-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US10836.
 XX
 PR 30-MAY-2000; 2000US-206132P.
 PR 29-AUG-2000; 2000US-228716P.
 XX
 PA (CURA-) CURAGEN CORP.
 PA Shinkets RA, Leach MD;
 PI Shinkets RA, Leach MD;
 XX
 DR WPI; 2002-106308/14.
 DR N-PSDB; ABN23167.
 XX
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders -
 XX
 PS Disclosure; SEQ ID 14812; 1037pp; English.
 XX
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX

RESULT 14
 AAY17499
 ID AAY17499 standard; Protein; 1525 AA.
 AC AAY17499;
 XX
 XX 04-AUG-1999 (first entry)
 DT
 XX Human Silt-1 protein.
 DE
 XX Human; Silt-1; Robo; modulation; identification; interaction.
 KW
 XX Homo sapiens.
 OS
 XX MO9925831-A2.
 PN
 XX 27-MAY-1999.
 PD
 XX 13-NOV-1998; 98MO-US24245.
 PF
 XX 07-APR-1998; 98US-0081057.
 PR 14-NOV-1997; 97US-0065544.
 XX
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Brose K, Goodman C, Kid T, Tessier-Lavigne M;
 PI
 XX WPI; 1999-347476/29.
 XX DR N-PSDB; AAX76161.
 XX
 XX Human Silt polypeptide and related nucleic acids
 PS Disclosure; Page 19-21; 34pp; English.
 XX
 XX The present sequence is the human Silt-1 protein. The present invention
 CC also describes a method for identifying agents which modulate the
 CC interaction of Robo and a Robo ligand comprising: combining a Robo
 CC polypeptide, a Silt polypeptide and a candidate agent under conditions
 CC where the Robo and Silt polypeptides normally (but for the presence of
 CC the agent) engage in a first interaction, where the Silt polypeptide
 CC specifically binds, activated or inhibits the activation of the Robo
 CC polypeptide and determining a second interaction of the Robo and Silt
 CC polypeptides in the presence of the agent, where a difference between
 CC the first and second interactions indicates that the agent modulates the
 CC interaction of the Robo and Silt polypeptides; and a method to modulate
 CC the interaction of Robo and a Robo ligand. The method is useful for
 CC screening for Robo (roundabout) modulators and Robo:Silt complexes are
 CC useful for regulating various cell functions, especially of neuronal
 CC cells.
 XX
 XX Sequence 1525 AA;
 SQ
 Query Match. 14.7%; Score 88; DB 20; Length 1525;
 Best Local Similarity 25.0%; Pred. No. 3;
 Matches 33; Conservative 28; Mismatches 63; Indels 8; Gaps 4;
 OY 1 LGAAVAFKQLPEPAALAAANKRVONILKKADALGEVNESLIQODE-EKALYAAAGIOP 59
 DB 544 LEATGIFKKLPQLRKINFSNNKTTDIEGAFEGASGVNEILLTSNRLNVQHKMKFGLE- 602
 OY 60 KIAAFAVEGNEFRTALSE-----LASVKPOVDAFDFGVMMVMAEDAAVKONRLNLRLAEQ 114
 DB 603 SLKTLMLRSNRITCVGNDSFGLSSVR-LLSLYDNOITTVAPGAFDTLHSLSTLNLAMP 661
 OY 115 MNAVADIALGE 126
 DB 662 FNCNCYLAWLGE 673

RESULT 15
 AAY27145
 ID AAY27145 standard; protein; 1529 AA.

XX
 AC AAY27145;
 XX
 XX 15-SEP-1999 (first entry)
 DT
 XX Human silt-2 protein (Seq ID NO: 12 of JP11164690).
 DE
 XX Vertebrate-derived protein; silt protein; diagnosis; cancer; nerve;
 KW muscle; endocrine system.
 XX
 XX Homo sapiens.
 OS
 XX JP11164690-A.
 PN
 XX 22-JUN-1999.
 PD
 XX 05-DEC-1997; 97JP-0335435.
 PF
 XX 05-DEC-1997; 97JP-0335435.
 PR
 XX (ASAH) ASAH KASEI KOGYO KK.
 PA
 XX WPI; 1999-411830/35.
 DR N-PSDB; AAX89162.
 XX
 XX New vertebrate silt protein - useful for diagnosis and treatment of
 PT cancers in nerves, muscle and endocrine system
 PS Disclosure; Page 58-63; 102pp; Japanese.
 XX
 XX The invention relates to a vertebrate-derived protein containing an
 CC amino acid sequence shown in AAY27137 and AAY27139. The
 CC vertebrate-derived protein has at least 55 % homology to one of sequences
 CC shown in AAY27141-127143, and has silt protein-like activity. The
 CC vertebrate silt proteins encoding nucleic acid sequences have at least
 CC 60% homology to nucleic acid sequences AAX89161-163. The
 CC vertebrate-derived proteins can be produced recombinantly by transforming
 CC host cells with expression vectors comprising the encoding nucleic acids.
 CC The proteins of the invention are for diagnosing and treating cancer of
 CC the nerves, muscle and/or endocrine system.
 XX
 XX Sequence 1529 AA;
 SQ
 Query Match. 14.7%; Score 88; DB 20; Length 1529;
 Best Local Similarity 25.0%; Pred. No. 3;
 Matches 33; Conservative 28; Mismatches 63; Indels 8; Gaps 4;
 OY 1 LGAAVAFKQLPEPAALAAANKRVONILKKADALGEVNESLIQODE-EKALYAAAGIOP 59
 DB 548 LEATGIFKKLPQLRKINFSNNKTTDIEGAFEGASGVNEILLTSNRLNVQHKMKFGLE- 606
 OY 60 KIAAFAVEGNEFRTALSE-----LASVKPOVDAFDFGVMMVMAEDAAVKONRLNLRLAEQ 114
 DB 607 SLKTLMLRSNRITCVGNDSFGLSSVR-LLSLYDNOITTVAPGAFDTLHSLSTLNLAMP 665
 OY 115 MNAVADIALGE 126
 DB 666 FNCNCYLAWLGE 677
 Search completed: December 2, 2002, 11:58:57
 Job time : 15.078 secs

Db 61 IAAVAEGNFRALSELASVYKPOVDAFFDGVVMAEDAAVKONRLNLRLAEQNNAVAD 120
QY 121 IALLGE 126
Db 121 IALLGE 126

RESULT 2

US-08-683-426-2
; Sequence 2, Application US/08683426
; Patent No. 5703367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,426
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-426-2

Query Match 100.0%; Score 598; DB 1: Length 126;

Best Local Similarity 100.0%; Pred. No. 3.7e-57; Mismatches 0; Indels 0; Gaps 0;

Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LOAVAVFKQLPEAAALAAANKRVONLLKKADAALGEVNESLLQODEEKALYAAAGLOPK 60
Db 1 LOAVAVFKQLPEAAALAAANKRVONLLKKADAALGEVNESLLQODEEKALYAAAGLOPK 60
QY 61 IAAVAEGNFRALSELASVYKPOVDAFFDGVVMAEDAAVKONRLNLRLAEQNNAVAD 120
Db 61 IAAVAEGNFRALSELASVYKPOVDAFFDGVVMAEDAAVKONRLNLRLAEQNNAVAD 120
QY 121 IALLGE 126
Db 121 IALLGE 126

RESULT 3

US-08-683-458-2
; Sequence 2, Application US/08683458

; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-458-2

Query Match 100.0%; Score 598; DB 1: Length 126;

Best Local Similarity 100.0%; Pred. No. 3.7e-57; Mismatches 0; Indels 0; Gaps 0;

Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LOAVAVFKQLPEAAALAAANKRVONLLKKADAALGEVNESLLQODEEKALYAAAGLOPK 60
Db 1 LOAVAVFKQLPEAAALAAANKRVONLLKKADAALGEVNESLLQODEEKALYAAAGLOPK 60
QY 61 IAAVAEGNFRALSELASVYKPOVDAFFDGVVMAEDAAVKONRLNLRLAEQNNAVAD 120
Db 61 IAAVAEGNFRALSELASVYKPOVDAFFDGVVMAEDAAVKONRLNLRLAEQNNAVAD 120
QY 121 IALLGE 126
Db 121 IALLGE 126

RESULT 4

US-08-878-360-2
; Sequence 2, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey

COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/683,426
FILING DATE:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-2

Query Match 100.0%; Score 598; DB 2; Length 126;
Best Local Similarity 100.0%; Pred. No. 3.7e-57;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQAVAVFKQLPEAAALAAANKRVONLLKKADALGEVNESLLOODEKALYAAAGLOPK 60
DB 1 LQAVAVFKQLPEAAALAAANKRVONLLKKADALGEVNESLLOODEKALYAAAGLOPK 60
QY 61 IAAVAEGNFRRLSLASVKKPVDAFPDGVYVMAEDAANKRNLNRLAQMNAVAD 120
DB 61 IAAVAEGNFRRLSLASVKKPVDAFPDGVYVMAEDAANKRNLNRLAQMNAVAD 120
QY 121 IALIGE 126
DB 121 IALIGE 126

RESULT 5
US-08-478-140B-2
Sequence 2, Application US/08478140B
Patent No. 6127153
GENERAL INFORMATION:
APPLICANT: JOHNSON, KARL F.
APPLICANT: ROTH, STEPHEN
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,140B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-140B-2

Query Match 100.0%; Score 598; DB 3; Length 126;
Best Local Similarity 100.0%; Pred. No. 3.7e-57;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQAVAVFKQLPEAAALAAANKRVONLLKKADALGEVNESLLOODEKALYAAAGLOPK 60
DB 1 LQAVAVFKQLPEAAALAAANKRVONLLKKADALGEVNESLLOODEKALYAAAGLOPK 60
QY 61 IAAVAEGNFRRLSLASVKKPVDAFPDGVYVMAEDAANKRNLNRLAQMNAVAD 120
DB 61 IAAVAEGNFRRLSLASVKKPVDAFPDGVYVMAEDAANKRNLNRLAQMNAVAD 120
QY 121 IALIGE 126
DB 121 IALIGE 126

RESULT 6
US-09-333-412-2
Sequence 2, Application US/09333412
Patent No. 6342382
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-JUN-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
type: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-333-412-2

Query Match	100.0%	Score 598;	DB 4;	Length 126;
Best Local Similarity	100.0%;	Pred. No. 3	7e-57;	
Matches 126;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	LQAAVAVKQLP	PEAAALAAANKRVONILKKDAALGEVNESLQODEEKALYAAAOGLQPK	60
Db	1	LQAAVAVEQJLP	PEAAALAAANKRVONILKKDAALGEVNESLQODEEKALYAAAOGLQPK	60
Qy	61	IAAAVABEGNFTAL	SELASVYPQVDAFEDGYMVAEDAAVKONLNLNRLAEOMNAVAD	120
Db	61	IAAAVABEGNFTAL	SELASVYPQVDAFEDGYMVAEDAAVKONLNLNRLAEOMNAVAD	120
Qy	121	IALIGE	126	
Db	121	IALIGE	126	

RESULT 7
US-09-338-943-2

GENERAL INFORMATION:
APPLICANT: JOHNSON, KARL F.
APPLICANT: ROTH, STEPHEN
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTTRANSFERASE AND GENE ENCODING A
TITLE OF INVENTION: POLYGLYCOSYLTTRANSFERASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/338,943
?

```

4.

MOLECULE TYPE: protein
US-09-338-943-2

Query Match	100.0%;	Score 598;	DB 4;	Length 126;
Best Local Similarity	100.0%;	Pred. No. 3.7e-57;		
Matches 126;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	LOAAVVEFQJLPEAAALAAANKRRVONLLKKADPAALGEYNESHL00DEEKALYAAAQGLQPK	60
Db	1	LOAAVVEFQJLPEAAALAAANKRRVONLLKKADPAALGEYNESHL00DEEKALYAAAQGLQPK	60
QY	61	IAAAVAEGNFPTALSETLASVAPÖYDAFEFGVMVAEDPAAYKÖNRLNLTNRLAEQMAVAD	120
Db	61	IAAAVAEGNFPTALSETLASVAPÖYDAFEFGVMVAEDPAAYKÖNRLNLTNRLAEQMAVAD	120
QY	121	IALTGE	126
Db	121	IALTGE	126

RESULT 8
US-08-844-085-2

```

1  GENERAL INFORMATION:
2  APPLICANT:  Lawlor, Elizabeth
3  TITLE OF INVENTION:  No. 5756330e1 Compounds
4  NUMBER OF SEQUENCES:  12
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE:  SmitKline Beecham Corporation
7  STREET:  709 Swedeland Road
8  CITY:  King of Prussia
9  STATE:  PA

```

```

1      COMPUTER READABLE FORM:
2      MEDIUM TYPE: Diskette
3      COMPUTER: IBM Compatible
4      OPERATING SYSTEM: DOS
5      SOFTWARE: FASTSOQ for Windows Version 2.0
6      CURRENT APPLICATION DATA:
7      APPLICATION NUMBER: US/08/844.085
8      FILING DATE: 18-APR-1997

```

```

; INFORMATION FOR SEQ ID NO: 2
; SEQUENCE CHARACTERISTICS:
; LENGTH: 678 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-844-085-2

```

Query Match	26.98;	Score 161;	DB 1;	Length 678;
Best Local Similarity	35.58;	Pred. No. 3.5e-09;		
Matches	43;	Conservative	21;	Mismatches 37;
			Indels	20;
			Gaps	3

```

QY      6 VFFQLEPAALAAANK-----RVQNLKKADALGEVNESILOODEKALYAA 53
      | | | | | : | | | | | : | : | : | | | | | |
Db      553 VADMLAASHLYEVSKKEEFKPSVESLSKAFNLAKAE-GVATYDSALFENDQEKALAEA 61
QY      54 AGLQLQKRIAAVAEGNFRLALSELASVYKPYQDAFFDGVMAAEDAAYKQNRNLNLNRLAE 113

```


;; CURRENT FILING DATE: 1998-06-23
;; NUMBER OF SEQ ID NOS: 10
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 2
;; LENGTH: 534
;; TYPE: PRF
;; ORGANISM: Streptococcus equi
US-09-103-664A-2

Query Match 12.9%; Score 77; DB 4; Length 534;
Best Local Similarity 24.7%; Pred. No. 2.8;
Matches 37; Conservative 21; Mismatches 40; Indels 52; Gaps 6;

QY 2 QAVAVFKQLPE--AAALAAANKRVONTL-----LKKADA 32
DB 260 KAFVYKSLADKLAKTAEAEKIMENVSGLRVESAKREMAQKLAETDQLTADAKADA 319
QY 33 ALGEVNESL--LQODEKALYAAAGLOPKTAAVAEGNFRTALSELASVRFVAFDQ 90
DB 320 ELAANDPTIASLQTELEKA-----KTELAVSERLIESGKREIAELQKQD----- 364
QY 91 VVWVEDAAVKONRLNLRLAEQMAVAD 120
DB 365 ----ASDKALVESQANVAE--LEKQKAASD 388

RESULT 13

US-08-080-255-7
; Sequence 7, Application US/08080255
; Patent No. 5487970
; GENERAL INFORMATION:
; APPLICANT: Rowley, Janet D.
; APPLICANT: Diaz, Manuel O.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND
; TITLE OF INVENTION: TRANSLOCATIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/080.255
; FILING DATE: 19930617
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-080-255-7

Query Match 12.5%; Score 75; DB 1; Length 1400;
Best Local Similarity 32.6%; Pred. No. 17;
Matches 29; Conservative 13; Mismatches 25; Indels 22; Gaps 4;

QY 5 AVFKQLEPAALAAANKRVONTLKKADALGEV--NESLQODEKALYAAAGLO---- 58
DB 227 SIGSMLAQADKLPMTDKRVASLLKRAKAQCKIEKSKSLKQTDQPK-----AQGQSDSS 281
QY 59 -----PKI-----AAVAEGNFRTALSE 76
DB 282 ETSVGRPRIKHVCRRRAVAALGRKRAVFPD 310

RESULT 14
US-08-465-713-7
; Sequence 7, Application US/08465713
; Patent No. 6121419
; GENERAL INFORMATION:
; APPLICANT: Rowley, Janet D.
; APPLICANT: Diaz, Manuel O.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND
; TITLE OF INVENTION: TRANSLOCATIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465.713
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/080.255
; FILING DATE: 17 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-465-713-7

Query Match 12.5%; Score 75; DB 3; Length 1400;
Best Local Similarity 32.6%; Pred. No. 17;
Matches 29; Conservative 13; Mismatches 25; Indels 22; Gaps 4;

QY 5 AVFKQLEPAALAAANKRVONTLKKADALGEV--NESLQODEKALYAAAGLO---- 58
DB 227 SIGSMLAQADKLPMTDKRVASLLKRAKAQCKIEKSKSLKQTDQPK-----AQGQSDSS 281
QY 59 -----PKI-----AAVAEGNFRTALSE 76
DB 282 ETSVGRPRIKHVCRRRAVAALGRKRAVFPD 310

RESULT 15
PCT-US93-05857-7
; Sequence 7, Application PC/TUS9305857
; GENERAL INFORMATION:
; APPLICANT: Board of Regents

APPLICANT: The University of Texas System
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: GENE REARRANGEMENTS AND TRANSLOCATIONS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05857
FILING DATE: 19930617
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/900,689
FILING DATE: 17/06/92
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: ARCD:072/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 320-7200
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-05857-7

Query Match 12.5%; Score 75; DB 5; Length 1400;
Best Local Similarity 32.6%; Pred. No. 17;
Matches 29; Conservative 13; Mismatches 25; Indels 22; Gaps 4;
QY 5 AVFKQLPEAAALAAANKRYQNLTKKADALGEV--NESLIQODEEKALYAAAGLQ----- 58
DB 227 STGSMAGQADKLPMTDKRVASILKKAKAQLCIKESKSLKQTDQPK-----AQGESDSS 281
QY 59 -----PKI-----AAVAEGNFRTALSE 76
DB 282 ETSVRCGRPIKHYCRAAVALGRRAVFPD 310

Search completed: December 2, 2002, 12:03:12
Job time : 6.69005 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 12:00:15 ; Search time 2.52541 Seconds
(without alignments)
794.504 Million cell updates/sec

Title: US-10-007-267-2

Perfect score: 598
Sequence: 1 LQAVAVFKQLPEAAALAAAN.....LINRLAEQMAVADIALIGE 126

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCF_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCFUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	598	100.0	126	US-10-007-267-2	Sequence 2, Appl1
2	163	27.3	989	US-09-815-242-4897	Sequence 4897, Ap
3	77.5	13.0	764	US-09-815-242-5143	Sequence 5143, Ap
4	75.5	12.6	271	US-09-847-539A-10	Sequence 10, Appl
5	75.5	12.6	1289	US-09-712-363-259	Sequence 259, Appl
6	75	12.5	905	US-09-905-963-58	Sequence 58, Appl
7	74	12.4	2186	US-09-815-242-12913	Sequence 12913, A
8	73.5	12.3	400	US-09-881-752A-206	Sequence 206, App
9	73.5	12.3	550	US-09-815-242-10499	Sequence 10499, A
10	73	12.2	2025	US-09-815-242-5703	Sequence 5703, Ap
11	73	12.2	3158	US-09-815-242-12611	Sequence 12611, A
12	72	12.0	393	US-09-815-242-11856	Sequence 11856, A
13	71.5	12.0	2437	US-09-815-242-5834	Sequence 5834, Ap
14	71.5	12.0	6281	US-09-815-242-12996	Sequence 12996, A
15	71	11.9	259	US-09-847-539A-8	Sequence 8, Appl
16	70	11.7	2434	US-09-815-242-5835	Sequence 5835, Ap
17	70	11.7	2478	US-09-815-242-5816	Sequence 5816, Ap
18	70	11.7	2478	US-09-815-242-12967	Sequence 12967, A
19	69.5	11.6	878	US-10-108-605-325	Sequence 325, App

20	69.5	11.6	1043	10	US-09-946-805-4	Sequence 4, Appl1
21	69	11.5	635	10	US-09-815-242-11209	Sequence 11209, A
22	69	11.5	1111	10	US-09-756-071B-15	Sequence 15, Appl
23	69	11.5	1193	10	US-09-756-071B-13	Sequence 13, Appl
24	69	11.5	1215	10	US-09-815-242-5908	Sequence 5908, Ap
25	69	11.5	1269	10	US-09-815-242-13113	Sequence 13113, A
26	69	11.5	2368	10	US-09-815-242-5635	Sequence 5635, Ap
27	69	11.5	2368	10	US-09-815-242-12389	Sequence 12389, A
28	68.5	11.5	355	10	US-09-741-669-435	Sequence 435, App
29	68.5	11.5	605	10	US-09-841-132-574	Sequence 574, App
30	68	11.4	234	10	US-09-939-960-517	Sequence 517, App
31	68	11.4	861	10	US-09-815-242-14088	Sequence 14088, A
32	68	11.4	1523	12	US-10-052-586-290	Sequence 290, App
33	67.5	11.3	266	12	US-10-067-989-3	Sequence 3, Appl1
34	66.5	11.1	592	9	US-10-047-593-2	Sequence 2, Appl1
35	66.5	11.1	592	9	US-10-047-593-4	Sequence 4, Appl1
36	66.5	11.1	874	10	US-09-815-242-11127	Sequence 11127, A
37	66.5	11.1	891	10	US-09-815-242-10147	Sequence 10147, A
38	66.5	11.1	903	10	US-09-905-963-56	Sequence 56, Appl
39	66	11.0	892	10	US-09-815-242-13843	Sequence 13843, A
40	65.5	11.0	577	10	US-09-815-242-13752	Sequence 13752, A
41	65.5	11.0	651	10	US-09-841-132-405	Sequence 405, App
42	65.5	11.0	651	12	US-10-007-693-78	Sequence 78, Appl
43	65.5	11.0	874	10	US-09-815-242-11809	Sequence 11809, A
44	65	10.9	303	10	US-09-925-297-740	Sequence 740, App
45	65	10.9	891	10	US-09-272-809-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-10-007-267-2
Sequence 2, Application US/10007267
Patent No. US20020127682A1
GENERAL INFORMATION:
APPLICANT: Gottschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-007-267-2

Query Match      100.0%; Score 598; DB 12; Length 126;
Best Local Similarity 100.0%; Pred. No. 8.2e-54;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQAAVAFKQLEPAALAAANKRVONLTKKADALGCVNESILQODEEKALYAAAGLQPK 60
D 1 LQAAVAFKQLEPAALAAANKRVONLTKKADALGCVNESILQODEEKALYAAAGLQPK 60
QY 61 IAAVAEGNFRTALSELASVKKPOVDAFPDGVYVMAEDAAYKONRLNLRLAEQMNAYAD 120
D 61 IAAVAEGNFRTALSELASVKKPOVDAFPDGVYVMAEDAAYKONRLNLRLAEQMNAYAD 120
QY 121 IALGGE 126
D 121 IALGGE 126

RESULT 2
US-09-815-242-4897
; Sequence 4897, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4897
; LENGTH: 989
; TYPE: PRP
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(989)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-4897

Query Match      27.3%; Score 163; DB 10; Length 989;
Best Local Similarity 31.5%; Pred. No. 1.1e-08;
Matches 46; Conservative 19; Mismatches 55; Indels 26; Gaps 3;

QY 1 LQAAVAFKQLEPAALAAANKRVONLTKKADALGCVNESILQODEEKALYAAAGLQPK 39
D 848 IDAVVSAEQADLSKLPASANIILKSPEDODFKPSMEALTRVINLAKGQELLGTGTEGID 907
```

```

QY 40 -SLIODEEKALYAAAGLQPKIAAANAEGNFRTALSELASVKKPOVDAFPDGVYVMAEDA 98
D 908 PSLEFNKAEKELYAVNDLSAFAFRTIAENYEA---LVNRLPIDAYFNETVWVWEDE 963
QY 99 AVKONRLNLRLAEQMNAYADIAL 124
D 964 KVKONRLKQIMQIMKMASTIASIDL 989

RESULT 3
US-09-815-242-5143
; Sequence 5143, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5143
; LENGTH: 764
; TYPE: PRP
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5143

Query Match      13.0%; Score 77.5; DB 10; Length 764;
Best Local Similarity 30.2%; Pred. No. 3.4;
Matches 42; Conservative 17; Mismatches 53; Indels 27; Gaps 7;

QY 2 QAAVAFKQLEPAALAAANKRVONLTKKADALGCVNESILQODEEKALYAAAGLQPK 58
D 209 QAAQVTVQULGADFNALTTTPATK-VQQRKAYADASLSQVSTQV-----SGAASATQANQ 262
QY 59 PKIAAANAEGN-----FRT--ALSELASVKKPOVDAFPDGVYVMAEDAAYKQ 102
D 263 VKTRAQAQQAQNSQISSQONLISAFAFNQJALAAQASGEAGVTFANQVQVINEATAFST 322
QY 103 NRNLNLRLAEQMNAYADI 121
D 323 ARKQALAOFA--ANAAAEV 339

RESULT 4
US-09-847-539A-10
; Sequence 10, Application US/09847539A
; Patent No. US20020061306A1
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
```



```
APPLICANT: Rassmussen, Magnus
TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
FILE REFERENCE: 100084.415US / N.75312B
CURRENT APPLICATION NUMBER: US/09/847,539A
CURRENT FILING DATE: 2001-05-01
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 271
TYPE: PRP
ORGANISM: Streptococcus pyogenes
US-09-847-539A-10
```

```
Query Match          12.6%; Score 75.5; DB 10; Length 271;
Best Local Similarity 27.5%; Pred. No. 1.4;
Matches 33; Conservative 16; Mismatches 40; Indels 31; Gaps 5;
```

```
OY 11 PEAAALAAANKRVQNLKKA-----DALGEVNESLIQODEEKLAAAGLOPK 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 28 PEKALRNEERAIDELKQAIEDKEATTAIEAASDALEALADQ-----ADALQSE 78
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 61 IAAVAEGFRRLASLAVKPOVDAFFDGVVMAEDAAVKQNRLLNRLAEQNNAVAD 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 79 EAAVQSDN--AASDALEALADQADA-----LQSEAAVQSD---NAAGDALEALAD 126
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 5
US-09-712-363-259
Sequence 259, Application US/09712363
GENERAL INFORMATION:
APPLICANT: Eisenberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712,363
CURRENT FILING DATE: 2000-11-13
PRIOR APPLICATION NUMBER: PCT/US00/02246
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: 60/179,531
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/117,844
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: 60/118,206,
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: 60/126,593
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 259
LENGTH: 1289
TYPE: PRP
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-259
```

```
Query Match          12.6%; Score 75.5; DB 9; Length 1289;
Best Local Similarity 25.0%; Pred. No. 11;
Matches 31; Conservative 21; Mismatches 65; Indels 7; Gaps 2;
```

```
OY 10 LPEAALAAANKRVQNLKKA-----AALGEVNESLIQODEEKLAAAGLOPKIAA 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 245 IEBAAGVLRHRRKREKALRKLDTMAANLRLDTLTRELROLPLGRQAEAAQRAAIAQA 304
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
OY 64 AVAEGNFRRLASLAVKPOVDAFFDGVVMAEDAAVKQNRLLN-RLAEQNNAVADIA 122
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 305 DLBDALRLRLAADDLVSRRAEREAIVFQAEAAAMREHDEAAARLAVASBELAAHSAVAELS 364
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 123 LIGE 126
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 365 TRAE 368
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 6
US-09-905-983-58
Sequence 58, Application US/09905983
Patent No. US20020045591A1
GENERAL INFORMATION:
APPLICANT: Geiger, Benjamin
APPLICANT: Ben-Ze'ev, Avri
APPLICANT: Sadot, Elnat
TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
FILE REFERENCE: 01/22326
CURRENT APPLICATION NUMBER: US/09/905,983
CURRENT FILING DATE: 2001-09-28
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn version 3.1
SEQ ID NO 58
LENGTH: 905
TYPE: PRP
ORGANISM: Mus musculus
US-09-905-983-58
```

```
Query Match          12.5%; Score 75; DB 10; Length 905;
Best Local Similarity 24.0%; Pred. No. 7.7;
Matches 30; Conservative 21; Mismatches 56; Indels 18; Gaps 4;
```

```
OY 3 AAVFKQLEPEAALAAANKRVQNLKKAADALGEVNESLIQODEKA-----LYAAAG 56
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 223 ASQAFRLHPDVATRANRDYVFKQVQDEAIGSAQAQATSPTEAKGHTGIGELAAALNE 282
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 57 LQKI---AAAVEGNFRRLASLAVKPOVDAFFDGVVMAEDAAVKQNRLLNRLAE 113
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 283 FDNKILDPMTSEARFRSLER-----RLSEITSGAALMAUDSCTRDNR---REEMVA 333
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 7
US-09-815-242-12913
Sequence 12913, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
```

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12913
; LENGTH: 2186
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match
Best Local Similarity 25.8%; Score 74; DB 10; Length 2186;
Matches 32; Conservative 20; Mismatches 52; Indels 20; Gaps 6;

QY 1 LQAVAVFK-QLPEPAALAAANKRVONLKKADAALGEVNESLQODEKALYA--AAQGL 57
Db 2000 LKAITALKQDEELKARTADVDVLRKRNVALSDI-EAVITEKESLRLIDINIAQOT 2058
QY 58 QPKIAAAAEENFTALSELASVKRQVDAF-DGVVMAEDAAVAKONRLNLNRLAEQMN 116
Db 2059 YAKFKAIATP-----EQLAKKVILIDQYVADGNRMIDEDAT-----LNDIKQHTQ 2103
QY 117 AVAD 120
Db 2104 FIVD 2107

RESULT 8
US-09-881-752A-206
; Sequence 206, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 206
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-206

Query Match
Best Local Similarity 12.3%; Score 73.5; DB 10; Length 400;
Matches 30; Conservative 16; Mismatches 44; Indels 17; Gaps 4;

QY 24 ONLKKADAALGEVNESLQ-----ODEKALYAAAGLQ---PKIAAVALGGR 71
Db 128 QNLAASNDVVLQAVFNLHQSITLKSMSQLSQEKEKELNTQALKNSIQKISSIIDOKTR 187
QY 72 TALSIELASVKPOVDAFDGVVMAEDAAVAKONRLNLNRLAEQMN 118
Db 188 EV--TLKSITKTEODKL--ILSMQKDYAIYQNLTLLEKRONLNL 229

RESULT 9
US-09-815-242-10499
; Sequence 10499, Application US/09815242
; Patent No. US20020061569A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10499
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10499

Query Match
Best Local Similarity 12.3%; Score 73.5; DB 10; Length 550;
Matches 31; Conservative 21; Mismatches 43; Indels 41; Gaps 3;

QY 8 KOLPEAALAAANKRVONLKKADA-----ALGEVNESLQODEKALYA 52
Db 27 KXDEKATEKTSVKVYLNLMENSEIGMSDIFTQDEASINAGSNVFEGLYQDEKDLIP 86
QY 53 AAQGLQPKTIAA-----AAEGNFRALSELASVKRQVDAF----- 88
Db 87 AAKKEPEISEDGKRYTIKLREDGKWSNGDAVTADEFVAMRKLANPKQANVFFLLBST 146
QY 89 --DGVVMAEDAAVAKO 102
Db 147 ILNGTAITKEEKAPPE 162

RESULT 10
US-09-815-242-5703
; Sequence 5703, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```

```
;
; TYPE: prt
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12611
Query Match          12.2%; Score 73; DB 10; Length 2025;
Best Local Similarity 25.9%; Pred. No. 36;
Matches 29; Conservative 20; Mismatches 55; Indels 8; Gaps 2;

QY 4 VAVFKQLPEPAALAAANKRVONLKKRADAALGEVNSLLQODEKALYAAAGLQPKIA 63
DB 113 IAVKQDLTEAQAQLDQMLNTLQOSIADKDATRASAVYVNAEPNKKQAYDEAVQNAESIING 172

QY 64 ----AAVEGNFRTALSELASVSPQVDAPFDGVVMAEDAAVKONRLNL 111
DB 173 LNNPTINKGNVSSATQAVTTSKNGL----DGVRLAQDKQTAGNSLNHLDQL 220

RESULT 11
US-09-815-242-12611
; Sequence 12611, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 12611
; LENGTH: 3158
```

```
;
; TYPE: prt
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12611
Query Match          12.2%; Score 73; DB 10; Length 3158;
Best Local Similarity 25.9%; Pred. No. 64;
Matches 29; Conservative 20; Mismatches 55; Indels 8; Gaps 2;

QY 4 VAVFKQLPEPAALAAANKRVONLKKRADAALGEVNSLLQODEKALYAAAGLQPKIA 63
DB 1244 IAVKQDLTEAQAQLDQMLNTLQOSIADKDATRASAVYVNAEPNKKQAYDEAVQNAESIING 1303

QY 64 ----AAVEGNFRTALSELASVSPQVDAPFDGVVMAEDAAVKONRLNL 111
DB 1304 LNNPTINKGNVSSATQAVTTSKNGL----DGVRLAQDKQTAGNSLNHLDQL 1351

RESULT 12
US-09-815-242-11856
; Sequence 11856, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11856
; TYPE: prt
; ORGANISM: Pseudomonas aeruginosa
; US-09-815-242-11856
Query Match          12.0%; Score 72; DB 10; Length 393;
Best Local Similarity 24.5%; Pred. No. 5.1;
Matches 34; Conservative 16; Mismatches 43; Indels 46; Gaps 5;

QY 31 DAALGEVNSLLQ-----DEKALYAAAGLQPKIAAAYAEVGNFR----- 71
DB 154 DYHMGITAEINLEAKYGLTEEDDAFAAAS--OOKAYIAIEGGEFRDEITPIQVPOKRGEP 211

QY 72 -----TALSELASVSP-----QVDAFEDGVVMAEDAAVKONRLNL- 107
DB 212 LQFDIDDEQPRAGTYEALKKLPAPFKKDSYTAGNASSLNDGAAAYVLMSAAKALGLIP 271

QY 108 -LNRLAEQNNAAVADIALLG 125
DB 272 VLARIASVAGVDPAIMG 290
```

```
RESULT 13
US-09-815-242-5834
; Sequence 5834, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5834
; LENGTH: 2437
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5834

Query Match      12.0%; Score 71.5; DB 10; Length 2437;
Best Local Similarity 26.0%; Pred. No. 64;
Matches 34; Conservative 22; Mismatches 62; Indels 13; Gaps 4;

QY 5 AVEKQPEAAALAAANKRYONLLKKADALGEVNESLQODEEKALYAAOGLQPKIAA- 63
DB 575 AVKQDTEQAQALDQMLDALQOSIADKDATRASSAYVNAEPNKKQSYDEAVQNAESIAGL 634
QY 64 --AAVEGNFRTALSELASVVKPOVDAFEDGVVMAEDAAYVKONRLNLNRL--AEQM--- 115
DB 635 NNPTINKGNVSSATQAVISSKNAL---DGVRLAQDKQTAGNSLNHLDLQLPAPAOQALE 690
QY 116 NAVADIALIGE 126
DB 691 NOINNAITRGE 701

RESULT 14
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 6281
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996

Query Match      12.0%; Score 71.5; DB 10; Length 6281;
Best Local Similarity 26.0%; Pred. No. 2,3e+02;
Matches 34; Conservative 22; Mismatches 62; Indels 13; Gaps 4;

QY 5 AVEKQPEAAALAAANKRYONLLKKADALGEVNESLQODEEKALYAAOGLQPKIAA- 63
DB 4417 AVKQDTEQAQALDQMLDALQOSIADKDATRASSAYVNAEPNKKQSYDEAVQNAESIAGL 4476
QY 64 --AAVEGNFRTALSELASVVKPOVDAFEDGVVMAEDAAYVKONRLNLNRL--AEQM--- 115
DB 4477 NNPTINKGNVSSATQAVISSKNAL---DGVRLAQDKQTAGNSLNHLDLQLPAPAOQALE 4532
QY 116 NAVADIALIGE 126
DB 4533 NOINNAITRGE 4543

RESULT 15
US-09-847-539A-8
; Sequence 8, Application US/09847539A
; Patent No. US20020061306A1
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN
; FILE REFERENCE: 100084.41505 / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-8

Query Match      11.9%; Score 71; DB 10; Length 259;
Best Local Similarity 27.5%; Pred. No. 3.7;
Matches 33; Conservative 17; Mismatches 42; Indels 28; Gaps 5;

QY 11 PEAALAAANKRYONLLKRA-----DALGEVNESLQODEEKALYAAOGLQPK 60
DB 45 PEKLAIRNERERAIDEIKKQIETKEATTAIEAASSALDELALDQ-----ADALQSE 95
QY 61 IAAVAEGNFRTALSELASVVKPOVDAF--FDGVVMAEDAAYVKONRLNLNRLAEQMNAY 118
DB 96 EAAVQSDN--AASDALDELALDQTDALQSEEAAYVVAADNNA-----SDFLDELALDQTDAL 148
```

```
FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 6281
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996
```

```
Query Match      12.0%; Score 71.5; DB 10; Length 6281;
Best Local Similarity 26.0%; Pred. No. 2.3e+02;
Matches 34; Conservative 22; Mismatches 62; Indels 13; Gaps 4;

QY 5 AVEKQPEAAALAAANKRYONLLKKADALGEVNESLQODEEKALYAAOGLQPKIAA- 63
DB 4417 AVKQDTEQAQALDQMLDALQOSIADKDATRASSAYVNAEPNKKQSYDEAVQNAESIAGL 4476
QY 64 --AAVEGNFRTALSELASVVKPOVDAFEDGVVMAEDAAYVKONRLNLNRL--AEQM--- 115
DB 4477 NNPTINKGNVSSATQAVISSKNAL---DGVRLAQDKQTAGNSLNHLDLQLPAPAOQALE 4532
QY 116 NAVADIALIGE 126
DB 4533 NOINNAITRGE 4543
```

```
RESULT 15
US-09-847-539A-8
; Sequence 8, Application US/09847539A
; Patent No. US20020061306A1
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN
; FILE REFERENCE: 100084.41505 / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-8
```

```
Query Match      11.9%; Score 71; DB 10; Length 259;
Best Local Similarity 27.5%; Pred. No. 3.7;
Matches 33; Conservative 17; Mismatches 42; Indels 28; Gaps 5;

QY 11 PEAALAAANKRYONLLKRA-----DALGEVNESLQODEEKALYAAOGLQPK 60
DB 45 PEKLAIRNERERAIDEIKKQIETKEATTAIEAASSALDELALDQ-----ADALQSE 95
QY 61 IAAVAEGNFRTALSELASVVKPOVDAF--FDGVVMAEDAAYVKONRLNLNRLAEQMNAY 118
DB 96 EAAVQSDN--AASDALDELALDQTDALQSEEAAYVVAADNNA-----SDFLDELALDQTDAL 148
```

Search completed: December 2, 2002, 12:15:58
Job time : 5.52541 secs

THIS PAGE BLANK (USPTO)


```
Query Match          35.2%; Score 210.5; DB 2; Length 722;
Best Local Similarity 41.0%; Pred. No. 7.2e-10;
Matches 50; Conservative 23; Mismatches 48; Indels 1; Gaps 1;

QY 1 LQAVAVFKQLPEAAALAAANKRVONLL---KKADALGCVNSLLOODEKALYAAAG 59
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 596 LDIAGRFALPEEALAVAKKRIKRLKRIKIPASVDTLDDQPAESCLVALLEGVIT 655
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 60 KIIAAVAEGNFRFALSELASVKPQVDAFFDGVVWMAEDAIVKONRLNLRLAEOMNVA 119
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 656 DTFSAIDCKNIVSVLCLARLPRIPEFDPKVVYNDENLMLRANRLALLORLEGIICVA 715
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 120 DI 121
   |||:|
Db 716 AI 717

RESULT 13
B97437
glycyl-tRNA synthetase, beta chain mmb1930 [imported] - Agrobacterium tumefaciens (strain C)
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: B97437
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollem, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: B97437
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-717 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK86451.1; PID:g15155593; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C1144
A:Map position: circular chromosome

Query Match          31.6%; Score 189; DB 2; Length 717;
Best Local Similarity 38.1%; Pred. No. 4.4e-08;
Matches 48; Conservative 20; Mismatches 54; Indels 4; Gaps 1;

QY 1 LQAVAVFKQLPEAAALAAANKRVONLL---KKADALGCVNSLLOODEKALYAAAG 56
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 588 VEAITAFITGEGDNRNLGAKKRAATOLLAAEKKGTVAADVSEELKLDKAKALYAIKT 647
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 57 LQPKIAAAVAEGNFRFALSELASVKPQVDAFFDGVVWMAEDAIVKONRLNLRLAEOMN 116
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 648 ASADAKAAVGEDEFRSAMQALSTLRAPVDKFFEDVLVNDDEDAIRANRLALLKAIREATG 707
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 117 AVADIA 122
   |||:|
Db 708 TVADFS 713

RESULT 14
AF2655
glycyl-tRNA synthetase, beta chain glys [imported] - Agrobacterium tumefaciens (strain C)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AF2655
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.; Gerage, G.; Gillet, W.; Grant, C.; Gunthner, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan, J.; Karp, P.; Romero, P.; Zhang, S. Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Teo, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.; Ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF2655
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-717 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41660.1; PID:g17739002; GSPDB:GN00186
```

```
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: glys
A:Map position: circular chromosome

Query Match          31.6%; Score 189; DB 2; Length 717;
Best Local Similarity 38.1%; Pred. No. 4.4e-08;
Matches 48; Conservative 20; Mismatches 54; Indels 4; Gaps 1;

QY 1 LQAVAVFKQLPEAAALAAANKRVONLL---KKADALGCVNSLLOODEKALYAAAG 56
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 588 VEAITAFITGEGDNRNLGAKKRAATOLLAAEKKGTVAADVSEELKLDKAKALYAIKT 647
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 57 LQPKIAAAVAEGNFRFALSELASVKPQVDAFFDGVVWMAEDAIVKONRLNLRLAEOMN 116
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 648 ASADAKAAVGEDEFRSAMQALSTLRAPVDKFFEDVLVNDDEDAIRANRLALLKAIREATG 707
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 117 AVADIA 122
   |||:|
Db 708 TVADFS 713

RESULT 15
AC3443
glycine-tRNA ligase (EC 6.1.1.14) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: AC3443
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov, J.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AC3443
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-780 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL52710.1; PID:g17983538; GSPDB:GN00190
C:Genetics:
A:Gene: BME11529
A:Map position: I
C:Superfamily: glycine-tRNA ligase beta chain
C:Keywords: ligase

Query Match          28.9%; Score 173; DB 2; Length 780;
Best Local Similarity 35.7%; Pred. No. 1e-06;
Matches 45; Conservative 22; Mismatches 55; Indels 4; Gaps 1;

QY 1 LQAVAVFKQLPEAAALAAANKRVONLL---KKADALGCVNSLLOODEKALYAAAG 56
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 651 LEALIVFINEVDGKNLLAGKRAANITLAAEKKGTVAADVADLRLAEAEKALFDVTL 710
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 57 LQPKIAAAVAEGNFRFALSELASVKPQVDAFFDGVVWMAEDAIVKONRLNLRLAEOMN 116
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 711 ASREVEGAIRAEDEFGMLALAKLRGVDIFFPKVLCVNDEDGWVRANRLALLDQIRATG 770
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 117 AVADIA 122
   |||:|
Db 771 KVADFS 776

Search completed: December 2, 2002, 12:00:05
Job time : 6.59198 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 3.06657 Seconds
(without alignments)
1704.189 Million cell updates/sec

Title: US-10-007-267-2

Perfect score: 598

Sequence: 1 LQAAVAFKQLEPEAAALAAAN.....LTNRLAEQMAVADIALIGE 126

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	598	100.0	126	SYGB_NEIGO	O50945 neisseria g
2	294.5	49.2	688	SYGB_VIBCH	O9KW8 vibrio chol
3	278	46.5	285	SYGB_HAEAE	O30836 haemophilus
4	273	45.7	688	SYGB_HAEIN	P43822 haemophilus
5	248.5	41.6	688	SYGB_ECOLT	P00961 escherichia
6	238	39.8	284	SYGB_COXRU	P45651 coxiella bu
7	231.5	38.7	689	SYGB_PASMO	P57905 pasteurella
8	210.5	35.2	722	SYGB_XYLFA	O9PC26 xyloella fas
9	151	25.3	672	SYGB_THEMA	O9WY60 thermotoga
10	138.5	23.2	665	SYGB_RICPR	O9ZCB1 rickettsia
11	138	23.1	679	SYGB_BAGSU	O67898 bacillus su
12	137	22.9	664	SYGB_AQUAE	O9K48 aquifex aeo
13	137	22.9	693	SYGB_BACHD	O9K48 bacillus ha
14	131.5	22.0	722	SYGB_SYNY3	O53690 synechocyst
15	130.5	21.8	690	SYGB_BUCAI	P57235 buchnera ap
16	86.5	14.5	810	HPR4_HALSA	O48317 halobacteri
17	84.5	14.1	189	PAAD_NEIMA	O9W78 neisseria m
18	83.5	14.0	810	HPR4_HALNI	O9H84 halobacteri
19	82.5	13.8	579	G160_HUMAN	O08378 homo sapien
20	82.5	13.8	579	G160_HUMAN	O08378 homo sapien
21	82	13.7	500	MSP_DIRIM	P13392 dirifloraria
22	82	13.6	505	FLUB_SALAE	P52616 salmonella
23	81.5	13.6	879	MYSR_ONCYO	O02171 onchocerca
24	81.5	13.6	880	MYSR_ONCYO	O02171 onchocerca
25	80.5	13.5	189	PAAD_NEIMA	O01202 brugia mala
26	79.5	13.3	239	RNC_HELPY	O9Y1P4 neisseria m
27	79.5	13.3	647	NAFL_MOUSE	P56118 helicobacte
28	79.5	13.3	882	MISP_CAEEL	O9WU8 mus musculu
29	78.5	13.1	239	RNC_HELPY	P15057 caenorhabdi
30	77.5	13.0	385	YL90_MYCTU	O9ZLH2 helicobacte
31	77.5	13.0	636	NAFL_HUMAN	O10383 mycobacteri
32	77	12.9	488	FLIC_SALCH	O15025 homo sapien
33	77	12.9	493	FLIC_SALPA	P06176 salmonella

34	77	12.9	513	1	GRDC_EUBAC	P54935 eubacterium
35	77	12.9	701	1	SYGB_HELPY	P56454 helicobacte
36	76.5	12.8	572	1	Y062_DEIRA	O9TRM6 deinococcus
37	76.5	12.8	1115	1	TBC2_CHLRE	O8VXP3 chlamydomon
38	76.5	12.8	2477	1	SPCN_CHICK	P07751 gallus gall
39	76	12.7	701	1	SYGB_HELPY	O9ZKM9 helicobacte
40	75.5	12.6	1205	1	SMC_MYCTU	O10970 mycobacteri
41	75	12.5	196	1	EFTS_THETH	P43895 thermus the
42	75	12.5	953	1	CTN2_HUMAN	P26232 homo sapien
43	75	12.5	953	1	CTN2_HUMAN	O61301 mus musculu
44	75	12.5	1325	1	G160_MOUSE	P55937 mus musculu
45	75	12.5	1426	1	CUT2_MOUSE	P70298 mus musculu

ALIGNMENTS

RESULT 1	ID	SYGB_NEIGO	STANDARD:	PRT:	126 AA.
AC	O50945:				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase				
DE	beta chain) (Glyrs) (Fragment).				
GN	GLYS				
OS	Neisseria gonorrhoeae.				
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
OX	NCBI_TaxID=485;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=P62;				
RX	MEDLINE=95053752; PubMed=7964493;				
RA	Gotschlich E.C.;				
RT	"Genetic locus for the biosynthesis of the variable portion of				
RL	Neisseria gonorrhoeae lipooligosaccharide.";				
J. Exp. Med.	180:2181-2190(1994).				
CC	- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) -> AMP + diphosphate				
CC	+ glycyl-tRNA(Gly).				
CC	- SUBUNIT: Tetramer of two alpha and two beta chains				
CC	(BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: Cytoplasmic.				
CC	- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U14554; AAA68008.1; -				
KW	PROSITE; P550861; AA-TRNA_LIGASE; II_GLYAB; 1.				
DR	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.				
FT	NON_TER				
FT	SEQUENCE 126 AA; 13343 MW; C3913AD1A6591CF2 CRC64;				
QY	Query Match	100.0%; Score 598; DB 1; Length 126;			
	Best local similarity	100.0%; Pred. No. 2.9e-43;			
	Matches 126; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 LQAAVAFKQLEPEAAALAAANKRVONILTKRADAALGVESNLSLQODEKALYAAAGLOPK 60				
DB	1 LQAAVAFKQLEPEAAALAAANKRVONILTKRADAALGVESNLSLQODEKALYAAAGLOPK 60				
QY	61 LQAAVEGFRRLASLAVKPVDAFFGVVWMAEDAIVKONRLNLRLAEQMAVAD 120				
DB	61 LQAAVEGFRRLASLAVKPVDAFFGVVWMAEDAIVKONRLNLRLAEQMAVAD 120				
QY	121 IALIGE 126				


```

ID  SYGB_XYLFA      STANDARD:      PRT:      722 AA.
AC  Q9PC26;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (glycine--tRNA ligase
DE  beta chain) (GLYRS).
GN  GLYS OR XF1959.
OS  Xylella fastidiosa.
OC  Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC  Xylella.
OX  Xylfa.
RX  NCBI_TaxID=2371;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=9a5c;
RX  MEDLINE=20365717; PubMed=10910347;
RA  Simpson A.J.G., Reinech F.C., Arruda P., Abreu F.A., Acencio M.,
RA  Alvenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA  Barros M.H., Bonaccorsi E.D., Bordin S., Boye J.M., Briones M.R.S.,
RA  Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,
RA  Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA  Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA  Facincani A.F., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA  Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA  Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA  Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA  Krieger J.E., Kuramae E.E., Laigret F., Lampis M.R., Leite L.C.C.,
RA  Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA  Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA  Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA  Menck C.F.M., Miranca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA  Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA  Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA  de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Patris A.,
RA  Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA  Quagão R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA  de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA  da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA  da Silveira J.F., Silvestri M.L.Z., Silveira W.J. de Souza A.A.,
RA  de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA  Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA  Zago M.A., Zatz M., Zeldanis J., Zetudal J.C.;
RT  "The genome sequence of the plant pathogen xylella fastidiosa.";
RT  Nature 406:151-159(2000).
CC  -i- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
CC  + glycyl-tRNA(Gly).
CC  -i- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
CC  (BY SIMILARITY).
CC  -i- SUBCELLULAR LOCATION: Cytoplasmic.
CC  -i- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch)
CC  -----
DR  EMBL: AE004015; AAF84761.1; -
DR  InterPro: IPR002311; tRNA_synth_2f.
DR  Pfam: PF02092; tRNA_synth_2f; 1.
DR  PRINTS: PRO1045; TRNASYNTHGB.
DR  TIGRFAMs: TIGR00211; glys; 1.
DR  PROSITE: PS50861; AA.tRNA.LIGASE.II.GLYAB; 1.
KW  Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW  Complete proteome.
SQ  SEQUENCE 722 AA; 80149 MW; AB6D0E4712B889EE CRC64;

```

```

Query Match      35.2%; Score 210.5; DB 1; Length 722;
Best Local Similarity 41.0%; Pred. No. 2.8e-10;
Matches 50; Conservative 23; Mismatches 48; Indels 1; Gaps 1;

```

```

QY  1 LQAAVAFKQLPPEAAALAAANKRVONLLKRAADAL-GEVNESILQODEKATLYAAAGLOP 59
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  596 LQAIFPALPEAEALAAVANKRRIRLTKRAEIKRIPASVATLFDQPAEGSLVALEGVTT 655
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY  60 KIAAFAVEGNFRTAISSELSAVKRFQVDAFFDGVWMAEDAAYKONRLNLRLAEQNNAVA 119
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  656 DFRSALDCKNKYSVLTCRLARLPDPEFDKVMVNDENIMLRNRLALLQREGEYICQVA 715
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY  120 DI 121
   |
DB  716 AI 717
   |
RESULT 9
SYGB_THEMEA
ID  SYGB_THEMEA      STANDARD:      PRT:      672 AA.
AC  Q9WY60;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (glycine--tRNA ligase
DE  beta chain) (GLYRS).
GN  GLYS OR TM0217.
OS  Thermotoga maritima.
OC  Bacteria; Thermotogae; Thermotogales;
OC  Thermotogaceae; Thermotoga.
OX  NCBI_TaxID=2336;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=MSB8 / DSM 3109;
RX  MEDLINE=99287316; PubMed=10360571;
RA  Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA  Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA  McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA  Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA  Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA  Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT  "Evidence for lateral gene transfer between Archaea and Bacteria from
RT  genome sequence of Thermotoga maritima.";
RT  Nature 399:323-329(1999).
CC  -i- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
CC  + glycyl-tRNA(Gly).
CC  -i- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
CC  (BY SIMILARITY).
CC  -i- SUBCELLULAR LOCATION: Cytoplasmic.
CC  -i- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch)
CC  -----
DR  EMBL: AE001706; AAD35309.1; -
DR  TIGR: TM0217; -
DR  InterPro: IPR002311; tRNA_synth_2f.
DR  Pfam: PF02092; tRNA_synth_2f; 1.
DR  PRINTS: PRO1045; TRNASYNTHGB.
DR  TIGRFAMs: TIGR00211; glys; 1.
DR  PROSITE: PS50861; AA.tRNA.LIGASE.II.GLYAB; 1.
KW  Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW  Complete proteome.
SQ  SEQUENCE 672 AA; 77813 MW; 612B402052DDA29A CRC64;

```

```

Query Match      25.3%; Score 151; DB 1; Length 672;
Best Local Similarity 30.1%; Pred. No. 2.2e-05;
Matches 37; Conservative 26; Mismatches 58; Indels 2; Gaps 1;
QY  2 QAAVAVFKQLPPEAAALAAANKRVONLLKRAADALGEVNESLQODEKATLYAAAGLOP 61
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  546 EALQEISEKPEFDQDLFVGFEVYHNITKNHDSI-KFDGALFEKEERKIMNKFYEYKEKV 603

```



```

DR EMBL: Z99117; CAB14468.1; -.
DR Subtilist; BG11658; glys.
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f.1.
DR PRINTS: PRO1045; TRNASYNTHB.
DR TIGRFAMs: TIGR00211; glys.1.
DR PROSITE: PS50861; AA_tRNA_LIGASE_II_GLYAB.1.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 679 AA; 76232 MW; ACC96723A859D2BB CRC64;

Query Match      23.1%; Score 138; DB 1; Length 679;
Best Local Similarity 28.8%; Pred. No. 0.00027;
Matches 34; Conservative 31; Mismatches 45; Indels 8; Gaps 2;

OY 7 PROLEPAALAAANKRVONLKKADALGEVNSLLQODEEKLTYAAAGLOPKIAAAY 66
DB 567 FETMEALG-----RVISISKK--GVREDIQDPDEFENEYKLFDPAYGTAKENIQENFS 618
OY 67 EGNFRTALSELASVVPQVDAFFDGVVMAEDAAYKONRLNLRLAEQNNAAVADIAL 124
DB 619 KNDYEALALASLAAKEPIDAYFDHMTVADNESLKANRLAQNVSILADEIKSFANMNA 676

RESULT 12
SYGB_AOUAE STANDARD; PRT; 664 AA.
AC 067898:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase
DE beta chain) (GLYS).
GN GLYS OR AQ.2141.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5.
RX MEDLINE-9819666; PubMed-9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Gerlach D.E., Overbeek R., Sneed M.A., Keller M., Aubay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:353-358(1998).
CC -i- CATALYTIC ACTIVITY: ATP + glycine + tRNA(gly) = AMP + diphosphate
CC + glycyl-tRNA(gly).
CC -i- SUBUNIT: Tetramer of two alpha and two beta chains
CC (by similarity).
CC -i- SUBCELLULAR LOCATION: Cytoplasmic.
CC -i- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000775; AAC07870.1; -.
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f.1.
DR PRINTS: PRO1045; TRNASYNTHB.
DR TIGRFAMs: TIGR00211; glys.1.
DR PROSITE: PS50861; AA_tRNA_LIGASE_II_GLYAB.1.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 664 AA; 77523 MW; 7FEF7A8F552E3DB CRC64;

```

```

Query Match      22.9%; Score 137; DB 1; Length 664;
Best Local Similarity 33.6%; Pred. No. 0.00032;
Matches 37; Conservative 18; Mismatches 45; Indels 10; Gaps 3;

OY 11 PPAALAAANKRVONLKKADALGEVNSLLQODEEKLTYAAAGLOPKIAAAYEGNF 70
DB 556 PEPFSVVEAKKRVIRIIPK-DWKNEYEVDEKLSEERELYOKLTFFENK-----EL 606
OY 71 RTALSELASVVPQVDAFFDGVVMAEDAAYKONRLNLRLAEQNNAAVAD 120
DB 607 KSPL-ELLPLKEYIDKFDPNVKVAEDEKIRNNRISILKRVENLRTFFED 655

RESULT 13
SYGB_BACHD STANDARD; PRT; 693 AA.
AC 09KD48:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase
DE beta chain) (GLYS).
GN GLYS OR BH1371.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE-20512582; PubMed-11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhnara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -i- CATALYTIC ACTIVITY: ATP + glycine + tRNA(gly) = AMP + diphosphate
CC + glycyl-tRNA(gly).
CC -i- SUBUNIT: Tetramer of two alpha and two beta chains
CC (by similarity).
CC -i- SUBCELLULAR LOCATION: Cytoplasmic.
CC -i- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001511; BAB05090.1; -.
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f.1.
DR PRINTS: PRO1045; TRNASYNTHB.
DR TIGRFAMs: TIGR00211; glys.1.
DR PROSITE: PS50861; AA_tRNA_LIGASE_II_GLYAB.1.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 693 AA; 77833 MW; 30F9919712105754 CRC64;

Query Match      22.9%; Score 137; DB 1; Length 693;
Best Local Similarity 28.2%; Pred. No. 0.00034;
Matches 35; Conservative 26; Mismatches 35; Indels 10; Gaps 2;

OY 6 VRQ-----LPEAALAAANKRVONLKKADALGEVNSLLQODEEKLTYAAAG 56
DB 564 VPKKAKILVSKVNPDEFKELVGLSRVTNIAKAEKNVA-IPDLPKKEEERVLVEAYVQ 622
OY 57 LDPKIAAAYAEENFRTALSELASVVPQVDAFFDGVVMAEDAAYKONRLNLRLAEQNN 116
DB 623 TRDLVQALASGDSVAAYAALEDTTEPIHGYPEHVWVMDQVYIKENRLAHAAAGVIG 682

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 10.8332 Seconds
(without alignments)
2398.732 Million cell updates/sec

Title: US-10-007-267-2

Sequence: 1 LQAVAVFKQLPEALAAAN.....LLNRLAEQNNVADIALGGE 126

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	575	96.2	687	16	Q9JXO5
2	572	95.7	687	16	Q9JW67
3	408	68.2	87	2	Q51114
4	313	52.3	697	16	Q8Y213
5	256.5	42.9	684	16	Q917B8
6	252.5	42.2	689	16	Q8XDN7
7	241.5	40.4	689	16	Q8Z9W7
8	229.5	38.4	689	16	Q8ZL96
9	229.5	38.4	689	16	Q8Z2B3
10	223	37.3	199	2	Q9F730
11	193	32.3	40	2	Q932S6
12	189	31.6	717	16	Q8UHN6
13	174	29.1	718	16	Q986B5
14	173	28.9	780	16	Q8YFJ1
15	164	27.4	673	16	Q9CGK0
16	160	26.8	678	16	Q97PW6

17	156	26.1	674	16	Q9ZG11	Q9ZG11 rickettsia
18	154	25.8	679	16	Q99Y14	Q99Y14 streptococ
19	154	25.8	721	16	Q9ZRL2	Q9ZRL2 rhizobium m
20	140.5	23.5	686	16	Q8RH44	Q8RH44 fusobacteri
21	138	23.1	692	16	Q9A8L1	Q9A8L1 caulobacter
22	124.5	20.8	716	16	Q8YPT0	Q8YPT0 anabaena sp
23	122	20.4	1067	10	Q23150	Q23150 arabidopsis
24	121	20.2	664	16	Q9PN63	Q9PN63 campylobact
25	119	19.9	1067	10	Q9SU73	Q9SU73 arabidopsis
26	103.5	17.3	688	16	Q8Y7S4	Q8Y7S4 listeria mo
27	92.5	15.5	688	16	Q9ZBQ2	Q9ZBQ2 listeria in
28	91	15.2	523	13	Q8YPF4	Q8YPF4 anabaena sp
29	91	15.2	1512	13	Q9DE36	Q9DE36 brachydanio
30	88	14.7	417	16	Q9ZLJ6	Q9ZLJ6 rhizobium m
31	88	14.7	1521	4	Q9S710	Q9S710 homo sapien
32	88	14.7	1525	4	Q9Y507	Q9Y507 homo sapien
33	88	14.7	1529	4	Q9A813	Q9A813 homo sapien
34	87	14.5	796	11	Q9WVC1	Q9WVC1 rattus norv
35	86.5	14.5	1010	16	Q8YK94	Q8YK94 anabaena sp
36	86	14.4	1025	11	Q9Z166	Q9Z166 mus musculu
37	86	14.4	1521	11	Q9R1B9	Q9R1B9 mus musculu
38	85.5	14.3	869	5	Q9NDV9	Q9NDV9 anisakis sl
39	85	14.2	1530	13	Q9OWZ3	Q9OWZ3 xenopus lae
40	84.5	14.1	423	17	Q9HQE5	Q9HQE5 halobacteri
41	84	14.0	522	5	Q18266	Q18266 caenorhabdi
42	84	14.0	804	1	P71416	P71416 halobacteri
43	83	13.9	449	2	Q91UV1	Q91UV1 escherichia
44	83	13.9	599	16	Q9XOC6	Q9XOC6 vibrio chol
45	82.5	13.8	259	4	Q9NXN5	Q9NXN5 homo sapien

ALIGNMENTS

RESULT 1
Q9JXO5 PRELIMINARY: PRT: 687 AA.
ID Q9JXO5
AC Q9JXO5: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glycyl-tRNA synthetase, beta chain.
GN NMB1930.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Yamatevan J.,
RA Gill J., Scariato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rapunoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RU Science 287:1809-1815 (2000).
DR EMBL: AE002541; AAF42259.1; -.
DR TIGR: NMB1930; -.
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f; 1.
DR PRINTS: PR01045; TRNASYNTHGB.
DR TIGRFAMs: TIGR00211; glys: 1.
KW Aminoacyl-tRNA synthetase; Complete proteome.
SQ SEQUENCE 687 AA; 74573 MW; 616BCBDD76A3D4FB CRC64;

Query Match 96.2%; Score 575; DB 16; Length 687;
Best Local Similarity 96.0%; Pred. No. 5.6e-39;
Matches 121; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

OY 1 LOAAVAFKQLEPEAAALAAANKRVONLLKKADALGEVNESILLQODEEKALYAAAQGIOPK 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 562 LOAAVAFKQLEPEAAALAAANKRVONLLKKADALGEVNESILLQODEEKALYAAAQGIOPK 621
OY 61 IAAAVAEGNFRTALSELASVYPQVDAFPDGVMAEDAAVQONRLNLRLAEQOMNAVAD 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 622 IAAAVAEGNFRTALSELASVYPQVDAFPDGVMAEDAAVQONRLNLRLAEQOMNAVAD 681
OY 121 IALIGE 126
    ||||| |||||
DB 682 IALIGE 687

RESULT 2
O9JW67 PRELIMINARY: PRT: 687 AA.
AC O9JW67;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Putative glycyl-tRNA synthetase beta subunit (EC 6.1.1.14).
GN GLYS OR NMA0523.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=2022256; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels R., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
DR EMBL: AL162753; CAB83815.1; -.
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f.1.
DR PRINTS: PR01045; TRNASYNTHGB.
DR TIGRFAMs: TIGR00211; glys.1.
KW Aminoacyl-tRNA synthetase; Ligase; Complete proteome.
SQ SEQUENCE 687 AA; 74650 MW; 8C17811D5C93EBE3 CRC64;

Query Match 95.7%; Score 572; DB 16; Length 687;
Best Local Similarity 96.0%; Pred. No. 9.9e-39;
Matches 121; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 LOAAVAFKQLEPEAAALAAANKRVONLLKKADALGEVNESILLQODEEKALYAAAQGIOPK 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 562 LOAAVAFKQLEPEAAALAAANKRVONLLKKADALGEVNESILLQODEEKALYAAAQGIOPK 621
OY 61 IAAAVAEGNFRTALSELASVYPQVDAFPDGVMAEDAAVQONRLNLRLAEQOMNAVAD 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 622 IAAAVAEGNFRTALSELASVYPQVDAFPDGVMAEDAAVQONRLNLRLAEQOMNAVAD 681
OY 121 IALIGE 126
    ||||| |||||
DB 682 IALIGE 687

RESULT 3
O51114 PRELIMINARY: PRT: 87 AA.
AC O51114;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE glycyl tRNA synthetase (fragment).
GN GLYS.

```

```

OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58;
RX MEDLINE=96414473; PubMed=8817494;
RA Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;
RT "Molecular analysis of a locus for the biosynthesis and phase-variable
RT expression of the lacto-N-neotetraose terminal lipopolysaccharide
RT structure in Neisseria meningitidis."
RL Mol. Microbiol. 18:729-740(1995).
RN [2]
RP SEQUENCE OF 12-87 FROM N.A.
RC STRAIN=126E;
RA Jennings M.P., Der Ley P., Poolman J.T., Moxon E.R.;
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: U25839; AAC44083.1; -.
DR EMBL: U65788; AAB48383.1; -.
KW Aminoacyl-tRNA synthetase.
FT NON_TER 1 1
SQ SEQUENCE 87 AA; 9242 MW; 22990BBD6FA2C8E0 CRC64;

Query Match 68.2%; Score 408; DB 2; Length 87;
Best Local Similarity 97.7%; Pred. No. 2.2e-26;
Matches 85; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 40 SLLOODEEKALYAAAQGIOPKIAAAVAEGNFRTALSELASVYPQVDAFPDGVMAEDAA 99
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 SLLOODEEKALYAAAQGIOPKIAAAVAEGNFRTALSELASVYPQVDAFPDGVMAEDAA 60
OY 100 VKONRLNLRLAEQOMNAVADIALIGE 126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 VKONRLNLRLAEQOMNAVADIALIGE 87

RESULT 4
O8Y213 PRELIMINARY: PRT: 697 AA.
AC O8Y213;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Probable glycyl-tRNA synthetase beta chain protein (EC 6.1.1.14).
GN GLYS OR RSC0524 OR RSC0494.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattoilco L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Layte W., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL: AL646059; CAD14052.1; -.
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f.1.
DR PRINTS: PR01045; TRNASYNTHGB.
DR TIGRFAMs: TIGR00211; glys.1.
KW Ligase; Complete proteome.
SQ SEQUENCE 697 AA; 75257 MW; 15A133DAD177EBDE CRC64;

Query Match 52.3%; Score 313; DB 16; Length 697;
Best Local Similarity 53.2%; Pred. No. 1.5e-17;
Matches 66; Conservative 19; Mismatches 39; Indels 0; Gaps 0;

```

OY 1 LGAAVAFKQLPEAAALAAANKRVONLKKRADAALGEVNESLLQODEKALYAAAGLOPK 60
 DB 572 LBAVAFAPALPQAEALAAANKRITNLTKTDTIGSVQKRLREDAERLHQAVANSEPH 631
 OY 61 IAAAVAEGRFALSELASVVKPOVDAFPDGVVMAEDAAVKONRLNLRLAEQNNAVD 120
 DB 632 VADAAARGDFTTALKTALSLREAVDSFFPDGVVMAEDDTRALRDLALGELGLMNRVAD 691
 OY 121 IALL 124
 DB 692 ISKL 695

RESULT 5

O917B8 PRELIMINARY: PRT: 684 AA.

AC 0917B8: PRELIMINARY: PRT: 684 AA.
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE glycyL-tRNA synthetase beta chain.
 GN GLYS OR PA0008.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stoyer C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 Garber R.L., Goltzy L., Tolentino E., Westbrook-Medman S., Yuan Y.,
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.";
 RT Nature 406:959-964(2000).
 RL EMBL: AE004440; AAC03398.1; -;
 DR InterPro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f.1.
 DR PRINTS: PR01045; TRNASYNTHGB.
 DR TIGRFAMS: TIGR00211; glys.1.
 KM Aminoacyl-tRNA synthetase; Complete proteome.
 SQ SEQUENCE 684 AA; 73973 MW; 74144075010F0BB6 CRC64;

Query Match 42.9%; Score 256.5; DB 16; Length 684;
 Best Local Similarity 48.4%; Pred. No. 6.2e-13;
 Matches 61; Conservative 20; Mismatches 44; Indels 1; Gaps 1;

OY 1 LGAAVAFKQLPEAAALAAANKRVONLKKRADAALGEVNESLLQODEKALYAAAGLOPK 59
 DB 559 VOAVDAFRLPPEAAALAAANKRVSNILAKSDEVPNNVASLVEAKALGASVANAES 618
 OY 60 KTAANAAGEFNFTALSELASVVKPOVDAFPDGVVMAEDAAVKONRLNLRLAEQNNAVD 119
 DB 619 EVAAPLAADYRAALALALALREPDVTFADYAVVAVDDAAVRANRYALAKLGRSLGVA 678
 OY 120 DIALG 125
 DB 679 DISLIG 684

RESULT 6

O8XDN7 PRELIMINARY: PRT: 689 AA.

AC 08XDN7: PRELIMINARY: PRT: 689 AA.
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Glycine tRNA synthetase, beta subunit.
 GN GLYS OR 24983 OR ECS4442.

OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RT Nature 409:529-533(2001).
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohnishi E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RT DNA Res. 8:11-22(2001).
 RL EMBL: AE005582; AAG58707.1; -;
 DR EMBL: AP002565; BAB37865.1; -;
 DR InterPro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f.1.
 DR PRINTS: PR01045; TRNASYNTHGB.
 DR TIGRFAMS: TIGR00211; glys.1.
 KM Aminoacyl-tRNA synthetase; Complete proteome.
 SQ SEQUENCE 689 AA; 76813 MW; 3F9B939FEA3168CD CRC64;

Query Match 42.2%; Score 252.5; DB 16; Length 689;
 Best Local Similarity 48.0%; Pred. No. 1.3e-12;
 Matches 60; Conservative 21; Mismatches 43; Indels 1; Gaps 1;

OY 1 LGAAVAFKQLPEAAALAAANKRVONLKKRADAALGEVNESLLQODEKALYAAAGLOPK 59
 DB 564 MKAVSHFRTLEAAALAAANKRVSNILAKSDEVLSRVNASTLKEPEIKLAMQVVVLDD 623
 OY 60 KTAANAAGEFNFTALSELASVVKPOVDAFPDGVVMAEDAAVKONRLNLRLAEQNNAVD 119
 DB 624 KLEPYFAEGRYDALVELALREPDVTFADYAVVAVDDDELRLNRLMLEKLEFLRAYA 683
 OY 120 DIALG 124
 DB 684 DISL 688

RESULT 7

O8Z9W7 PRELIMINARY: PRT: 689 AA.

AC 08Z9W7: PRELIMINARY: PRT: 689 AA.
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE glycyL-tRNA synthetase beta subunit (EC 6.1.1.14).
 GN GLYS OR YPO4071.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CO-92 / BIOVAR ORIENTALIS;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebalnia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,

RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Katterishev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.,
 RT "genome sequence of *Yersinia pestis*, the causative agent of plague."
 RL Nature 413:523-527(2001).
 DR EMBL: AJ414160; CAC93524.1; -
 DR InterPro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f; 1.
 DR PRINTS: PR01045; TRNASYNTHGB.
 DR TIGRFAMs: TIGR00211; glys; 1.
 KW Aminoacyl-tRNA synthetase; Ligase; Complete proteome.
 SQ SEQUENCE 689 AA; 76204 MW; 715AF2AEED693937 CRC64;

Query Match 40.4%; Score 241.5; DB 16; Length 689;
 Best Local Similarity 46.4%; Pred. No. 1,le-11;
 Matches 58; Conservative 23; Mismatches 43; Indels 1; Gaps 1;

OY 1 LOAAVAFKQLEPAAALAAANKRVONLKKADAALGE-VNESLQODEKALYAAAGIOP 59
 Db 564 VKAVYFRTLLAAALAAANKRVSNILAKSTDTLDHVASILKPAELKLTFLVLVLD 623
 OY 60 KIAAFAEGNFTALSELASVKKPOVDAFEDGVVMAEDAAVKNRLNLRLAEQMANVA 119
 Db 624 QLEPYFAAGQIKREALVELALRETYDEFFESVMAEDAAVKNRLTLISKIRELFQYA 683
 OY 120 DIALL 124
 Db 684 DISL 688

RESULT 8

O82L96 PRELIMINARY; PRT; 689 AA.
 AC O82L96;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Glycine tRNA synthetase, beta subunit (EC 6.1.1.14).
 GN GLYS OR STM3655.
 OS *Salmonella typhimurium*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Salmonella*.
 OC NCBI_TaxID=602;
 RP [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 LT2.";
 RL Nature 413:852-856(2001).
 DR EMBL: AB008869; AAL22514.1; -
 DR InterPro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f; 1.
 DR PRINTS: PR01045; TRNASYNTHGB.
 DR TIGRFAMs: TIGR00211; glys; 1.
 KW Aminoacyl-tRNA synthetase; Ligase; Complete proteome.
 SQ SEQUENCE 689 AA; 76452 MW; 5734C31C5223123F CRC64;

Query Match 38.4%; Score 229.5; DB 16; Length 689;
 Best Local Similarity 44.8%; Pred. No. 1e-10; Indels 1; Gaps 1;
 Matches 56; Conservative 26; Mismatches 42; Indels 1; Gaps 1;

OY 1 LOAAVAFKQLEPAAALAAANKRVONLKKADAALGE-VNESLQODEKALYAAAGIOP 59
 Db 564 MKAVSHFRTLEASALAAANKRVSNILAKATEPLNDIVHASVLEKAEATFLAHLVYLVD 623
 OY 60 KIAAFAEGNFTALSELASVKKPOVDAFEDGVVMAEDAAVKNRLNLRLAEQMANVA 119

Db 624 KLOPYFADRGQELIELALRAVDEFFENVVMAEKDIRIRITLTLISKIRELFQYA 683
 OY 120 DIALL 124
 Db 684 DISL 688

RESULT 9

O82B3 PRELIMINARY; PRT; 689 AA.
 AC O82B3;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Glycine tRNA synthetase, beta subunit (EC 6.1.1.14).
 GN STY4144.
 OS *Salmonella typhi*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Salmonella*.
 OC NCBI_TaxID=601;
 RP [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL: AL627281; CAD07973.1; -
 DR InterPro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f; 1.
 DR PRINTS: PR01045; TRNASYNTHGB.
 DR TIGRFAMs: TIGR00211; glys; 1.
 KW Aminoacyl-tRNA synthetase; Ligase; Complete proteome.
 SQ SEQUENCE 689 AA; 76645 MW; A93CC239351028D9 CRC64;

Query Match 38.4%; Score 229.5; DB 16; Length 689;
 Best Local Similarity 44.8%; Pred. No. 1e-10; Indels 1; Gaps 1;
 Matches 56; Conservative 26; Mismatches 42; Indels 1; Gaps 1;

OY 1 LOAAVAFKQLEPAAALAAANKRVONLKKADAALGE-VNESLQODEKALYAAAGIOP 59
 Db 564 MKAVSHFRTLEASALAAANKRVSNILAKATEPLNDIVHASVLEKAEATFLAHLVYLVD 623
 OY 60 KIAAFAEGNFTALSELASVKKPOVDAFEDGVVMAEDAAVKNRLNLRLAEQMANVA 119
 Db 624 KLOPYFADRGQELIELALRAVDEFFENVVMAEKDIRIRITLTLISKIRELFQYA 683
 OY 120 DIALL 124
 Db 684 DISL 688

RESULT 10
 O9F730 PRELIMINARY; PRT; 199 AA.
 AC O9F730;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE Glys-like protein (Fragment).
 OS *Pseudomonas fluorescens*.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC *Pseudomonas*.
 OC NCBI_TaxID=294;


```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-F113;
RX MEDLINE=20476421; PubMed=11021923;
RA Laue B.E., Jiang Y., Chhabra S.R., Jacob S., Stewart G.S.A.B.,
RT "The biocontrol strain Pseudomonas fluorescens F113 produces the
RT rhizobium small bacteriocin, N-(3-hydroxy-7-cis-
RT tetradecenyl)homoserine lactone, via Hds, a putative novel N-
RT acylhomoserine lactone synthase."
RL Microbiology 146:2469-2480(2000).
DR EMBL: AF286536; ANG30824.1; -.
DR Interpro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f; 1.
FT NON_TER 1
SQ SEQUENCE 199 AA; 21900 MW; 464CD941AA21A7C1 CRC64;

Query Match 37.3%; Score 223; DB 2; Length 199;
Best Local Similarity 40.8%; Pred. No. 7.9e-11;
Matches 51; Conservative 26; Mismatches 48; Indels 0; Gaps 0;

QY 1 LOAAVAFKQLEPAALAAANKRVQNLKKAADALGEVNESLLOQDEKALYAAAGLQPK 60
Db 75 VOAVQAFKLEPEAAALAAANKRVSNLSKAEPSYKAVEARYFDNAAEFSNSAIQCAVYAS 134
QY 61 IAAVAEGNEFRALSELASVYKPOVDAFFDGVWMAEDAAVKONRLNLRLAEQMAVAD 120
Db 135 VOPLMEKRDYAEALRLAALREPVDAFFEAVMVNAQDEYRKNRYLALRLSLFLNIAD 194
QY 121 IALLG 125
Db 195 ISVLG 199

RESULT 11
Q932S6 PRELIMINARY; PRT; 40 AA.
ID Q932S6;
AC Q932S6;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE GLYS (Fragment).
GN GLYS.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M978; AND A1;
RX MEDLINE=21467954; PubMed=11583844;
RA Zhu P., Klutch M.J., Tsai C.-M.;
RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
RT Expression in Two 18-Immuno-type Strains of Neisseria meningitidis."
RL FEMS Microbiol. Lett. 203:173-177(2001).
DR EMBL: AF355193; AAL12837.1; -.
DR EMBL: AF355194; AAL12843.1; -.
FT NON_TER 1
SQ SEQUENCE 40 AA; 4392 MW; E41CA1A42B65A187 CRC64;

Query Match 32.3%; Score 193; DB 2; Length 40;
Best Local Similarity 100.0%; Pred. No. 3.3e-09;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 FFDGVWMAEDAAVKONRLNLRLAEQMAVADIALGE 126
Db 1 FFDGVWMAEDAAVKONRLNLRLAEQMAVADIALGE 40

RESULT 12
Q98UH6 PRELIMINARY; PRT; 717 AA.
ID Q98UH6;
AC Q98UH6;
DT 01-JUN-2002 (TReMBLrel. 21, Created)

```

```

DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE GLYC1-crna synthetase, beta chain.
GN GLYS OR ATU0644 OR AGR_C.1144.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. Sr.,
RA Chapman P., Clendenning J., Decherage G., Gillet W., Grant C.,
RA Kutayavir T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RT Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlet K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RT Science 294:2323-2328(2001).
DR EMBL: AE009032; AAL41660.1; -.
DR EMBL: AE007998; AAK86451.1; -.
KW Aminocycl-crna synthetase; Complete proteome.
SQ SEQUENCE 717 AA; 78713 MW; 94F26877B8EF048 CRC64;

Query Match 31.6%; Score 189; DB 16; Length 717;
Best Local Similarity 38.1%; Pred. No. 2.2e-07;
Matches 48; Conservative 20; Mismatches 54; Indels 4; Gaps 1;

QY 1 LOAAVAFKQLEPAALAAANKRVQNL----KKADAALGEVNESLLOQDEKALYAAAG 56
Db 588 VEALTAFTTGEDGRNLLAGKATQTLLAAEEKGTVAADVSEELKLDKALYAAIKT 647
QY 57 LOPRIAAVAEGNEFRALSELASVYKPOVDAFFDGVWMAEDAAVKONRLNLRLAEQMA 116
Db 648 ASADAAKAVGCEPFRSMQALSTIRAPVDFEVDVYNDEDAIRANRLALAIRATG 707
QY 117 AVADIA 122
Db 708 TVADFS 713

RESULT 13
Q986B5 PRELIMINARY; PRT; 718 AA.
ID Q986B5;
AC Q986B5;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE GLYC1-crna synthetase, beta chain.
GN MLR7435.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=MAF303099;
RA MEDLINE=21082930; PubMed=11214968;
RX Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Matanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL, AF003011; BAB53538.1; -
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f; 1.
DR PRINTS: PRO1045; TRNASYNTHGB.
DR TIGRFAMS: TIGR00211; glys; 1.
KW Aminocyl-tRNA synthetase; Complete proteome.
SQ SEQUENCE 718 AA; 78610 MW; 6664029CFAC669 CRC64;

Query Match 29.1%; Score 174; DB 16; Length 718;
Best Local Similarity 32.5%; Pred. No. 3.7e-06;
Matches 41; Conservative 26; Mismatches 55; Indels 4; Gaps 1;

OY 1 LOAVAVFKQLEPAALAAANKRVONLL-----KKADALGEVNESLLOODEKALYAAAG 56
Db 589 VEALGSFDTEDGKNLLGTRKRAANIILAEKKKTAVAKTVEPALFKENAEKSLFAAVNQ 648
OY 57 LOPKIAAIVAEGNFRTALSELASVKPOVDAPFDGVMVAEDAAYKONRLNLRLAEQMN 116
Db 649 AEKQAGEAIQNDPESAMALALSLAREPDSFPEGVLVNDDELVEVANRLALLTRIAATG 708
OY 117 AVADIA 122
Db 709 QVADFS 714

RESULT 14

OQ8YFJ1 PRELIMINARY; PRT; 780 AA.
AC 08YFJ1;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14).
GN BMEI1529.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA Delvecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykdis A., Resnik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyripides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL, AE009588; AAL52710.1; -
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f; 1.
DR TIGRFAMS: TIGR00211; glys; 1.
KW ligase; Complete proteome.
SQ SEQUENCE 780 AA; 85795 MW; 27947D756B21FE80 CRC64;

Query Match 28.9%; Score 173; DB 16; Length 780;
Best Local Similarity 35.7%; Pred. No. 4.9e-06;
Matches 45; Conservative 22; Mismatches 55; Indels 4; Gaps 1;

OY 1 LOAVAVFKQLEPAALAAANKRVONLL-----KKADALGEVNESLLOODEKALYAAAG 56

Db 651 LEALIVEINEVDGKNLLAGAKRANIIILAEKKGTQVADVADVPALRLREAEKALFDVTL 710
OY 57 LOPKIAAIVAEGNFRTALSELASVKPOVDAPFDGVMVAEDAAYKONRLNLRLAEQMN 116
Db 711 ASREVGAEIARDPENGAMALAKLRGPVDFEDKVLVNDGNGVANRLALLDQIRATG 770
OY 117 AVADIA 122
Db 771 KVADFS 776

RESULT 15

OQ9CGKO PRELIMINARY; PRT; 673 AA.
AC 09CGKO;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14).
GN GLT OR LI1096.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LI1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Boletín A., Winkler P., Mauger S., Jallón O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403";
RL Genome Res. 11:731-753(2001).
DR EMBL, AE006341; AAK05194.1; -
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f; 1.
DR PRINTS: PRO1045; TRNASYNTHGB.
DR TIGRFAMS: TIGR00211; glys; 1.
KW Aminocyl-tRNA synthetase; Complete proteome.
SQ SEQUENCE 673 AA; 75717 MW; 99CF078804DAE69E CRC64;

Query Match 27.4%; Score 164; DB 16; Length 673;
Best Local Similarity 38.0%; Pred. No. 2.2e-05;
Matches 38; Conservative 21; Mismatches 37; Indels 4; Gaps 2;

OY 25 NLKKADALGEVNESLLOODEKALYAAAGLQPKRIAAVAEGNFRTALSELASVKPOV 84
Db 575 NLVKA-KDIKEINPTLFEDAEALYNNVISLQNMWTYMPGEEKRAIVHSLA---PAI 630
OY 85 DAFDGVMAEDAAVAKONRLNLRLAEQMNADIALI 124
Db 631 ETFEESVMAEDLSVRDRIALLSEVALTSVMADFSII 670

Search completed: December 2, 2002, 12:02:13
Job time : 12.8232 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 36.1203 Seconds
(without alignments)
1283.801 Million cell updates/sec

Title: US-10-007-267-3
Perfect score: 1823
Sequence: 1 MQPLVSLICAVNVEKYFAQ.....RLIKNRQARSACKKEEI 348

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1823	100.0	348	17	AA191311
2	1820	99.8	348	18	AAW06576
3	1709	93.7	346	23	AAU72923
4	1049	57.5	337	17	AA191314
5	1046	57.4	337	18	AAW06579
6	254.5	14.0	274	22	AA191315
7	244	13.4	274	20	AAW89330
8	227	12.5	278	21	AAV68976
9	226.5	12.4	334	21	AAV68963
10	225	12.3	298	22	AA191313

11	217.5	11.9	324	18	AAW14078
12	217.5	11.9	324	18	AAW22177
13	214.5	11.8	316	21	AAV54072
14	214.5	11.8	316	21	AAV43774
15	211	11.6	332	22	AA191314
16	210	11.5	322	21	AAV68974
17	210	11.5	322	21	AAV68974
18	210	11.5	322	21	AAV68974
19	210	11.5	322	21	AAV68974
20	208.5	11.4	322	21	AAV68974
21	208.5	11.4	322	21	AAV68974
22	207.5	11.4	322	21	AAV68974
23	207.5	11.4	322	21	AAV68974
24	205	11.2	322	21	AAV68974
25	200.5	11.0	322	21	AAV68974
26	200	11.0	322	21	AAV68974
27	195	10.7	322	21	AAV68974
28	194.5	10.7	322	21	AAV68974
29	193.5	10.6	322	21	AAV68974
30	193.5	10.6	322	21	AAV68974
31	191.5	10.5	322	21	AAV68974
32	191.5	10.5	322	21	AAV68974
33	190.5	10.4	322	21	AAV68974
34	190	10.4	322	21	AAV68974
35	189.5	10.4	322	21	AAV68974
36	185.5	10.2	322	21	AAV68974
37	183	10.0	322	21	AAV68974
38	181	9.9	322	21	AAV68974
39	179	9.8	322	21	AAV68974
40	179	9.8	322	21	AAV68974
41	177	9.7	322	21	AAV68974
42	177	9.7	322	21	AAV68974
43	175	9.6	322	21	AAV68974
44	175	9.6	322	21	AAV68974
45	175	9.6	322	21	AAV68974

ALIGNMENTS

RESULT 1	AA191311	standard; Protein; 348 AA.
ID	AA191311	
AC	AA191311	
DT	09-JUL-1996	(first entry)
DE	N. gonorrhoeae glycosyltransferase LgtA.	
KW	Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus; vaccine.	
OS	Neisseria gonorrhoeae strain F62.	
PN	W09610086-A1.	
PD	04-APR-1996.	
PF	25-SEP-1995;	95WO-US12317.
PR	26-SEP-1994;	94US-0312387.
XX	(UYRQ) UNIV ROCKEFELLER.	
XX	Gotschlich EC;	
XX	WPI: 1996-200924/20.	
XX	N-PSDB; AAT14061.	
XX	Nucleic acids encoding glycosyl transferase(s) - used in the	
XX	diagnosis of infection with Neisseria and for the biosynthesis of	
XX	oligo:saccharide(s)	

XX Claim 8; Fig 2b; 81pp; English.
PS
CC 5 Glycosyltransferases (AAR91311-15) are products of the lgt locus
CC (AAT1061) of *Neisseria gonorrhoeae* strain F62. Glycosyltransferase
CC LgtA (AAR91311) can be obtd. by expression of the lgtA coding
CC sequence in recombinant host cells. A method for adding GalNAc
CC or GlcNAc betal-3 to Gal comprises contacting a reaction mixture
CC contg. activated GalNAc or GlcNAc to an acceptor moiety comprising
CC a Gal residue in the presence of LgtA. Oligosaccharides can be
CC produced that, when attached to non-toxic lipids, are useful for
CC *Neisseria* vaccine prepn. Blood group core oligosaccharides, and
CC mimics of lacto-N-neotriose, gangliosides and saccaride portions
CC of globoglycolipids can also be produced using the enzymes.
XX
XX Sequence 348 AA;

Query Match 100.0%; Score 1823; DB 17; Length 348;
Best Local Similarity 100.0%; Pred. No. 3.2e-184;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPLYSVLICAVNVEKYFAQSIAAVVNOTWRNLDTLIYDDGSTDTLAIAKDFQRDSRI 60
DB 1 MOPLYSVLICAVNVEKYFAQSIAAVVNOTWRNLDTLIYDDGSTDTLAIAKDFQRDSRI 60
QY 61 KILAOAONSGILIPSLNIGDELAKSGGGGEXIARTDADDIASPGMIKIVGEMEKDSRI 120
DB 61 KILAOAONSGILIPSLNIGDELAKSGGGGEXIARTDADDIASPGMIKIVGEMEKDSRI 120
QY 121 IAMGAMLEVLSEKDGKGNRLARHHKHKGIWKKPTRHEDIAAEPFGNPJHNTMIMRSYI 180
DB 121 IAMGAMLEVLSEKDGKGNRLARHHKHKGIWKKPTRHEDIAAEPFGNPJHNTMIMRSYI 180
QY 181 DGLRLDTFRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLAHANOVSSKHSVROHEIAQI 240
DB 181 DGLRLDTFRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLAHANOVSSKHSVROHEIAQI 240
QY 241 OKTARNDFLOSNGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
DB 241 OKTARNDFLOSNGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
QY 301 PSGAWLDFADGMRRLFTLROYFGILYRLIKNRROARSDSAGKEOEI 348
DB 301 PSGAWLDFADGMRRLFTLROYFGILYRLIKNRROARSDSAGKEOEI 348

RESULT 2
AAW06576
ID AAW06576 standard; Protein; 348 AA.
XX
AC AAW06576;

XX 21-MAR-1997 (first entry)
XX
DE *Neisseria polyglycosyltransferase*.
XX
XX *Polyglycosyltransferase; N-acetylglucosaminyl transferase;*
KW *N-acetylglucosaminyl transferase; lipo-oligosaccharide*.
XX
OS *Neisseria gonorrhoeae* ATCC 33084.
XX
XX
XX WO9640971-A1.
XX
XX 19-DEC-1996.
XX
XX 03-JUN-1996; 96WO-US08323.
XX
XX 07-JUN-1995; 95US-0478140.
XX
XX (NEOS-) NEOFSE TECHNOLOGIES INC.
XX
XX Bucala SL, Johnson KF, Roth S;
XX
XX

DR WPI: 1997-052351/05.
DR N-PSDB: AAT49230.
XX
XX Transfer of at least 2 saccharide units using
PT *polyglycosyltransferase* - isolated from *N. gonorrhoeae*, catalyses
PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
PT single galactose moiety
XX
XX
PS Disclosure: Fig 2A-C; 38pp; English.
XX
XX
XX A novel *polyglycosyltransferase* (PGTase) (AAW06576) from *Neisseria*
CC *gonorrhoeae* ATCC 33084 catalyses the stereospecific conjugation of
CC 2 specific activated saccharide units (e.g. UDP-GlcNAc, UDP-GalNAc,
CC UDP-Gal) to specific acceptors having a galactose moiety at a
CC non-reducing end. It is the first PGTase reported to be capable of
CC transfer of more than one different saccharide moiety. The PGTase
CC is encoded by nucleotides 445-1488 of a lipo-oligosaccharide gene
CC (AAT49230). It can be produced in transformed host cells and used in
CC oligosaccharide prodn.
XX
XX Sequence 348 AA;

Query Match 99.8%; Score 1820; DB 18; Length 348;
Best Local Similarity 99.7%; Pred. No. 6.6e-184;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPLYSVLICAVNVEKYFAQSIAAVVNOTWRNLDTLIYDDGSTDTLAIAKDFQRDSRI 60
DB 1 MOPLYSVLICAVNVEKYFAQSIAAVVNOTWRNLDTLIYDDGSTDTLAIAKDFQRDSRI 60
QY 61 KILAOAONSGILIPSLNIGDELAKSGGGGEXIARTDADDIASPGMIKIVGEMEKDSRI 120
DB 61 KILAOAONSGILIPSLNIGDELAKSGGGGEXIARTDADDIASPGMIKIVGEMEKDSRI 120
QY 121 IAMGAMLEVLSEKDGKGNRLARHHKHKGIWKKPTRHEDIAAEPFGNPJHNTMIMRSYI 180
DB 121 IAMGAMLEVLSEKDGKGNRLARHHKHKGIWKKPTRHEDIAAEPFGNPJHNTMIMRSYI 180
QY 181 DGLRLDTFRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLAHANOVSSKHSVROHEIAQI 240
DB 181 DGLRLDTFRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLAHANOVSSKHSVROHEIAQI 240
QY 241 OKTARNDFLOSNGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
DB 241 OKTARNDFLOSNGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
QY 301 PSGAWLDFADGMRRLFTLROYFGILYRLIKNRROARSDSAGKEOEI 348
DB 301 PSGAWLDFADGMRRLFTLROYFGILYRLIKNRROARSDSAGKEOEI 348

RESULT 3
AAU72923
ID AAU72923 standard; Protein; 346 AA.
XX
AC AAU72923;

XX 12-MAR-2002 (first entry)
XX
XX *Neisseria meningitidis* virulence protein #13.
DE
KW *Meningitidis* virulence; gene; antibacterial; vaccine; veterinary;
KW infection; Gram-negative bacteria; antimicrobial.
XX
XX *Neisseria meningitidis*.
XX
XX WO200185772-A2.
XX
XX 15-NOV-2001.
XX
XX 08-MAY-2001; 2001WO-GB02003.
XX
XX 08-MAY-2000; 2000GB-0011108.
XX
XX

```

XX (MICR-) MICROSCIENCE LTD.
PA
XX Tang C:
PI
XX WPI: 2002-066593/09.
DR N-PSDB: AAS97208.
XX
PT New peptide encoded by operon including virulence genes of Neisseria
PT meningitidis, useful as vaccine component for treating or preventing
PT meningitis and for identifying antimicrobial drug
XX
XX Claim 4; Page 60-61; 423pp; English.
XX
CC The invention relates to a peptide (I) encoded by an operon (II) of
CC Neisseria meningitidis including virulence genes, or a related molecule
CC having a 40% sequence similarity at the peptide or nucleotide level in a
CC Gram-negative bacterium, or its functional fragment, for therapeutic or
CC diagnostic use. (I) and (II) are useful in the manufacture of a
CC medicament for treating or preventing a condition (e.g., meningitis)
CC associated with infection by Neisseria or Gram-negative bacteria. The
CC product is useful for veterinary treatment and in a screening assay for
CC the identification of an antimicrobial drug. The vaccines have
CC prophylactic applications. AA072911-AA073014 represent N. meningitidis
CC virulence proteins of the invention.
XX
SQ Sequence 346 AA:
XX
Query Match 93.7%; Score 1709; DB 23; Length 346;
Best Local Similarity 94.0%; Pred. No. 3.8e-172;
Matches 327; Conservative 11; Mismatches 8; Indels 2; Gaps 1;
XX
QY 1 MOPVLVLCAYVVEKFFAOSLAAYVNOTWRNDLIIYDGSIDGTLATAKDQKRSRI 60
DB 1 LQPLVSLICAVNEKFFAOSLAAYVNOTWRNDLIIYDGSIDGTLATAKDQKRSRI 60
XX
QY 61 KILAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMKDRSI 120
DB 61 KILAQNSGLIPSLNIGDELAKS--GMGEYIARTDADDIAAPDWIEKIVGEMKDRSI 118
XX
QY 121 IAMGAVLEVLSEKDGKRLARHHKGIKKPTRHEDIAAFPEFGNPINNTMIMRSYI 180
DB 119 IAMGAVLEVLSEKDGKRLARHHKGIKKPTRHEDIAAFPEFGNPINNTMIMRSYI 178
XX
QY 181 DGLRDTEDMAEDYQFWYDYSKLGRLAYYPEALVKYRLHANOVSSKSVRHEIAOGI 240
DB 179 DGLRNTEDMAEDYQFWYDYSKLGRLAYYPEALVKYRLHANOVSSKSVRHEIAOGI 238
XX
QY 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDP 300
DB 239 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDP 298
XX
QY 301 PSGAWLDFADGMRRLFTLROYFGILYRLIKNRQARSDSAGKEOEI 348
DB 299 PAGAWLDFADGMRRLFTLROYFGILYRLIKNRQARSDSAGKEOEI 346
XX
RESULT 4
AAR91314
ID AAR91314 standard; Protein; 337 AA.
XX
AC AAR91314;
XX
DT 09-JUL-1996 (first entry)
XX
DE N. gonorrhoeae glycosyltransferase LgtD.
XX
KM Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
XX vaccine.
XX OS Neisseria gonorrhoeae strain F62.
XX
PN WO9610086-A1.

```

```

XX PD 04-APR-1996.
XX
XX PF 25-SEP-1995; 95WO-US12317.
XX
XX PR 26-SEP-1994; 94US-0312387.
XX
XX PA (UYRQ) UNIV ROCKEFFELLER.
XX
XX PI Gotschlich EC;
XX
XX DR WPI: 1996-200924/20.
XX N-PSDB: AAT14061.
XX
PT Nucleic acids encoding glycosyl transferase(s) - used in the
PT diagnosis of infection with Neisseria and for the biosynthesis of
PT oligo:saccharide(s)
XX
XX Claim 11; Fig 2e; 81pp; English.
XX
CC 5 Glycosyltransferases (AAR91311-15) are products of the lgt locus
CC (AAT14061) of Neisseria gonorrhoeae strain F62. Glycosyltransferase
CC LgtD (AAR91314) can be obt. by expression of the lgt coding
CC sequence in recombinant host cells. A method for adding GalNAc
CC or GlcNAc beta1-3 to Gal comprises contacting a reaction mixture
CC contg. activated GalNAc or GlcNAc to an acceptor moiety comprising a
CC Gal residue in the presence of LgtD. Oligosaccharides can be produced
CC that, when attached to non-toxic lipids, are useful for Neisseria
CC vaccine prepn. Blood group core oligosaccharides, and mimics of
CC lipo-N-neoteirase, gangliosides and saccharide portions of
CC globoglycolipids can also be produced using the enzymes.
XX
SQ Sequence 337 AA:
XX
Query Match 57.5%; Score 1049; DB 17; Length 337;
Best Local Similarity 64.5%; Pred. No. 2.9e-102;
Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;
XX
QY 1 MOPVLVLCAYVVEKFFAOSLAAYVNOTWRNDLIIYDGSIDGTLATAKDQKRSRI 60
DB 1 MOPVLVLCAYVVEKFFAOSLAAYVNOTWRNDLIIYDGSIDGTLATAKDQKRSRI 60
XX
QY 61 KILAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMKDRSI 120
DB 61 RIISNRRNIGFSLNIGDELAKS--GGGEYIARTDADDIASPGWIEKIVGEMKDRSI 118
XX
QY 121 IAMGAVLEVLSEKDGKRLARHHKGIKKPTRHEDIAAFPEFGNPINNTMIMRSYI 180
DB 119 IAMGAVLEVLSEKDGKRLARHHKGIKKPTRHEDIAAFPEFGNPINNTMIMRSYI 178
XX
QY 181 DGLRDTEDMAEDYQFWYDYSKLGRLAYYPEALVKYRLHANOVSSKSVRHEIAOGI 240
DB 179 DGLRDPAYIHAEDYKFWYEAAGKLGRLAYYPEALVKYRFHODQTSKKNLQDRRTAMKI 238
XX
QY 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDP 300
DB 239 KERIRAGYKKAAGIAGACACINGLLAKSTAYALYKALSGODIGCLRLTYEFTLSLEKI 298
XX
QY 301 PSGAWLDFADGMRRLFTLROYFGILYRLIK 332
DB 299 SLTDLDLDFLTRVYMRKLFAPYRKILKKMLR 330
XX
RESULT 5
AAW06579
ID AAW06579 standard; Protein; 337 AA.
XX
AC AAW06579;
XX
DT 21-MAR-1997 (first entry)
XX
DE lipo-oligosaccharide gene-encoded protein.
XX

```

KW	Polyglycosyltransferase; N-acetylglucosaminyl transferase;
KM	N-acetylglactosaminyl transferase; Lipo-oligosaccharide.
OS	Neisseria gonorrhoeae ATCC 33084.
XX	
XX	WO9640971-A1.
PN	
PD	19-DEC-1996.
XX	
PF	03-JUN-1996; 96MO-US08323.
PR	
PR	07-JUN-1995; 95US-0478140.
XX	
PA	(NEOS-) NEOSE TECHNOLOGIES INC.
PI	
PI	Buczala SL, Johnson KF, Roth S;
XX	
DR	WPI; 1997-052351/05.
DR	N-PSDB; AAT9230.
XX	
PT	Transfer of at least 2 saccharide units using
PT	polyglycosyltransferase - isolated from N. gonorrhoeae, catalyses
PT	the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
PT	single galactose moiety
XX	
XX	
XX	Disclosure; Fig 2E-F; 38pp; English.
XX	
CC	A lipooligosaccharide-encoding gene region (AAT9230) of Neisseria
CC	gonorrhoeae ATCC 33084. Includes coding sequences for 5 proteins
CC	(AAW06576-80), one of which (AAW06576) is a polyglycosyltransferase
CC	that catalyses the addition of GlcNAc and GalNAc disaccharides to
CC	a galactose moiety. The function of the other 4 proteins is not
XX	stated in the specification.
XX	
XX	Sequence 337 AA;
XX	

Query Match	57.4%;	Score 1046;	DB 18;	Length 337;
Best Local Similarity	64.2%;	Pred. No. 6.1e-102;		
Matches 213;	Conservative 34;	Mismatches 83;	Indels 2;	Gaps 1

Oy	1	MOPVSVLVCAYNNEKKEFAAGSLAVVMQNRNNDILLVDGSEFGTALAKDKQKRSRI	60
Oy	1	LOPLVSVLVCAYNNEKKEFAAGSLAVVMQNRNNDILLVDGSEFGTALAKDKQKRSRI	60
Oy	61	KILAQONSGILPSLNTGLDELAASGGGGGGETARTADDIASPGWLEKIVGEMEKDRST	120
Oy	61	RIISNPRLGFIASLNTGLDELAAS - GGEGGYARTADDIASPGWLEKIVGEMEKDRST	118
Oy	121	IAMGAMLEVLSEEEKDGNRLARHNHKGKIMKPRTHEDIAAFFPGPNDIHNNMTIMRSVY	180
Oy	119	IAMGAMLEVLSEEEKNNKSVTLAIAANGAIMPKPTHEHEDIVAFPPGNDIHNNMTIMRSVY	178
Oy	181	DGGLRTYTERMDAEDYQPMYIVSVSKLGRATYYPALVYKYLRLANOVSSKHSVROHETLOGI	240
Oy	179	DGGLNFEDATYIHAEDYKRWYAGSLKGRATYYPALVYKYLRLANOVSSKHSVROHETLOGI	238
Oy	241	OKTARNDFLOSGMGFETRDSLEKYQOTKAAALEPEKPLPEDEPERARFPLYOCFKRTDTP	300
Oy	239	KKEEINAGWYKMAAGIANGVADCLNTNGILKSTVALYELKALSGDDICLRLTYEYFLEKY	298
Oy	301	PSGAMLDFAADGRRKRLFTLQYIGILYRLIK	332
Oy	299	SLTDLDELDTDRVMKILFAADYQKRLIKKMLR	330

RESULT6	
AAG90151	
ID	AAG90151 standard; Protein; 274 AA
XX	
AC	AAG90151;
XX	
DT	26-SEP-2001 (first entry)
XX	

DE C glutamicum protein fragment SEQ ID NO: 3905.

KM Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis.

PF	18-DEC-2000;	2000EP-0127688.
XX		
PR	16-DEC-1999;	99JP-0377484.
PR	07-APR-2000;	2000JP-0159162.
PR	03-AUG-2000;	2000JP-0280988.

PA (KYOW) KYOWA HAKKO KOGYO KK.

PI Nakagawa S,	Mizoguchi H, Ando S, Hayashi M,	Ochiai K,	Yokoi H
PI Tateishi N,	Senoh A,	Ikeda M,	Ozaki A;

DR WPI; 2001-376931/40
DR N-PSDB; AAH65370.

PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX -
PS Claim 17; SEQ ID NO: 3905; 246bp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the *Corynebacterium Corynebacterium glutamicum*. These are useful for identifying the mutation point of a gene derived from a mutant of *Corynebacterium* bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from *Corynebacterium* bacterium, and identifying a homologue of a gene derived from *Corynebacterium* bacterium. *Corynebacterium* bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.

Sequence 274 AA.

Query Match	14.0%	Score 254.5	DB 22	Length 274
Best Local Similarity	30.0%	Try. NO. 3e-18		
Matches 80	Conservative 4%	Mismatches 91	Indels 47	Gaps 10

[illegible]

[illegible][illegible]

AC	AAT68963,
XX	
DT	30-MAY-2000 (first entry)
XX	
DE	Cps2k protein which has glycosyltransferase activity.
XX	
KM	Capsular gene cluster; serotype 2; polysaccharide biosynthesis;
KM	capsular component; antigen; regulation; chain length determination;
KW	complement-mediated opsonophagocytosis; serotype-specific detection;
KW	antigen; vaccine; Streptococcal disease; ORF 22; ORF 2Y; ORF 2Z;
KM	Cps2A; Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J;
KM	Cps2K; Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T.
XX	
OS	Streptococcus suis.
PN	WO200005378-A2.
XX	
PD	03-FEB-2000.
XX	
PF	19-JUL-1999; 99WO-NL00460.
XX	
PR	22-JUL-1998; 98EP-0202465.
PR	22-JUL-1998; 98EP-0202467.
XX	
PA	(DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
PI	Smith HE;
DR	WPI: 2000-195104/17.
N-PSDB:	AAZ60929.
PT	New nucleic acid containing the capsular gene cluster of Streptococcus
PT	suis, used for serotype-specific detection and to generate antigens or
XX	mutants for vaccination
PS	Disclosure; Fig 3; 144pp; English.
XX	
CC	The proteins AAY68950-69 are encoded by the capsular gene cluster of
CC	Streptococcus suis serotype 2. The genes in this cluster are involved
CC	in polysaccharide biosynthesis of capsular components and antigens. The
CC	proteins are involved in regulation (CpsA), chain length determination
CC	(CpsB, CpsC), export (CpsC), and biosynthesis (CpsE, CpsF, CpsG, CpsH,
CC	CpsI, CpsK). The capsule confers bacterium resistance to
CC	complement-mediated opsonophagocytosis. The gene cluster is used as a
CC	source of probes and primers for serotype-specific detection of S. suis
CC	and is also useful for recombinant production of the proteins. The
CC	proteins are then useful for producing antigens that can be used in
CC	vaccines, for controlling or eradicating a Streptococcal disease, in
CC	humans or animals, e.g. against S. suis in pigs.
XX	
S0	Sequence 334 AA;
Query Match:	12.4%; Score 226.5; DB 21; Length 334;
Best Local Similarity	22.7%; Pred. No. 3.8e-15;
Matches 84; Conservative	73; Mismatches 98; Indels 115; Gaps 15.
OY	5 VSVLLCAVNEKYFAQSIAAVVNQTWRNDLIIVDGSGTDGLTAIAKFQKRDSRIKILA 64 : : : : : : : : : : : : : : : : : : : :
Dd	4 ISIIPIYNVEDEYSKCINSITVNOYKHIELLVLDGSDNSEELCLAAVAKKDSRIRYT- 62 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY	65 QAONSGILPDLNIGIDELAKSGGGGEYTARTDADDIASPGNIETKIIVGEMEKDRSTIAMG 124 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd	63 KRENGELISDARNYGISR-AK-----GDYLAFIDSDDFIHSEFIORHEATERNALVVAVA 116 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY	125 AWLEVLSEKKGNRLAKRHNNKGIKKRPRIHNDIAAFPPFGCPRIINNNTIMIR-RSYIDGG 183 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd	117 GDHY-----DASGHFTADLPRTNDAVLSSGRNVCKKL 149 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY	184 LRYDER-----DMAEDYOF-----WYDSKGRLAUYDEALVVKXL 220 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd	150 LEADGHRVVAAMNKLYKKELLEDFEFPEKGKINHEDEYFYTRLLYLEKAIVAIVECLYYVD 209 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY	221 HANOV--SSKHSVRHEITAOGIOKTARNDFLOSMGFKTRFDLSLEVROTKAAUYELPEKDL 278 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db	210	RENSIIITSSMTDHRHCLE--FQNERMDFYESRG-----DKEL	246
Oy	279	PEDEFERARFLYOCFKRDTPPSGAMLDFAA-----DGRMRRLFTLLRQYFGILY	328
Db	247	-----LLECCR-----SFLFAVLFLGKYNHMLSKQCKL--LQTLFRIVY	285
Oy	329	RLIK-NRROA 337	
Db	286	KOLKONKRLA 295	
RESULT 10			
ID	AAB96313	standard; Protein; 298 AA.	
XX			
AC	AAB96313;		
XX			
DT	29-OCT-2001	(first entry)	
XX			
DE		putative glycosyltransferase, involved in cell wall biogenesis #1.	
XX			
KW		Hyperthermophilic archaeon; hyperthermophilic protein.	
XX			
OS		Pyrococcus abyssi.	
XX			
PN	FR2792651-A1.		
PD	27-OCT-2000.		
XX			
XX	21-APR-1999;	99FR-0005034.	
XX			
PR	21-APR-1999;	99FR-0005034.	
XX			
PA	(CNRS) CNRS CENT NAT RECH SCI.		
PA	(IFREMER) IFREMER INST FR RECH EXPL MER.		
XX			
PI	Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;		
PI	Querellou J, Weissenbach J, Saurin W, Hellig R;		
DR	WPI, 2001-126236/14.		
XX			
PT	New nucleotide sequences isolated from Pyrococcus abyssi encode		
PT	proteins useful in industry -		
XX			
PS			
CC	Claim 7; Pages 981-982; 1657pp; French.		
CC			
CC	The present invention relates to the genomic sequence of Pyrococcus		
CC	abyssi (see AAF84431 and AAH41223-7) and P. abyssi proteins. P. abyssi is		
CC	a hyperthermophilic archaeon, which is isolated from deep-sea		
CC	hydrothermal vents. The present sequence is one such P. abyssi protein.		
CC	The proteins of the present invention have various potential industrial		
CC	uses, since the proteins are stable at very high temperatures, some up to		
CC	110 degrees centigrade.		
CC	Note: This patent is in the same patent family as WO20065062, which		
CC	contains additional sequences as shown in AAB99132-AAB99143,		
CC	AAH5903-AAH75920 and AAG66436.		
XX			
XX			
Sequence	298 AA:		
Query Match	12.3%; Score 225; DB 22; Length 298;		
Best Local Similarity	26.5%; Pred. No. 4,6e-15;		
Matches	65; Conservative 52; Mismatches 96; Indels 32; Gaps 8		
Oy	2	OPVLVLCATVVEKTFPAOSLAAYVNOQRNIDILVDDGSDGTALAKDFOKRDSRIK	61
Db	3	REIVSVIIPTYNRANLRLRAIVSYLNOKKDELLIVDDASDNPVEVSEIE--DGRIR	60
Oy	62	ILAAQONSLIPSLINIGDELAKSGGGGEYIARIDADDIAFPCMTIEKIVGEME---KRR	118
Db	61	YIRLRKNSGGPIARNIGIKK-AK-----GRFALLDDDEWMLPHRLEVGVRRFENLGKEF	114
Oy	119	SIIAAGAMLEVIYSEKDKGNRLARHHRHGKIVMKPTRHEDIAAFPPGQNIHNNTMIMRS	178

[illegible][illegible]

```

Db      63 SVTSK-ENGGSNARNRQIKK-AR-----GSFTTFVSDDIYKDYLSHLVAGKSTSI 115
Qy      121 IAMGAWL-----EVLSEB-----KQGNRLARHHKHKIWK- 151
        :      :      :      :      :      :      :      :
Db      116 VCSKFLELDEKSGLLTKKAPKKKSEVYSIESIKILLLOONGDYLA-----VWGKLY 168
        :      :      :      :      :      :      :      :
Qy      152 PTRHEDIAAFPPFQNPINHNNTMIKRSVLDGGLYDTERPMABEYQWYVYSGKIGRAY 211
        :      :      :      :      :      :      :      :
Db      169 PVSEFETISF-----DEGKIYEDMGTTYLLKLKLAEEVF 202
        :      :      :      :      :      :      :      :
Qy      212 PEAL-VKRYRLHANV-SKSHSVROHEIAOGIQTARNDFLOSMGFKRFPISLE-YROTKA 268
        :      :      :      :      :      :      :      :
Db      203 LDAVDAYAVQRPNSIMNSFLLKLLDILEAMVE-MENDIL-----AQPNLLALYKNRA 255
        :      :      :      :      :      :      :      :
Qy      269 AA-----YELPEKDLPEDEFEARRRFLYQCFKKTDP 301
        :      :      :      :      :      :      :      :
Db      256 FAAEYKIFLELIPK-----EKEFEQAKOKOLMDHKKNNRRAP 290

```

AAW22172	RESULT 12
ID AAW22177	standard; Protein: 324 AA.
XX	
AC AAW22177;	
XX	
DT 29-AUG-1997	(first entry)
XX	
DE 5.thermophilus exopolysaccharide synthesis operon epsI gene product.	
XX	
KW Streptococcus thermophilus; operon: exopolysaccharide; synthesis;	
KW open reading frame; thickening; drink; liquid dessert; yoghurt; soup;	
KW sauce; ice cream; coffee cream; mayonnaise; glycosyl transferase.	
XX	
OS Streptococcus thermophilus.	
XX	
PN EP750043-A1.	
XX	
PD 27-DEC-1996.	
XX	
PF 28-DEC-1995;	95EP-0203663.
XX	
PR 20-JUN-1995;	95EP-0201669.
XX	
PA (NEST) SOC PROD NESTLE SA.	
XX	
PI Mollet B, Stinglele F;	
XX	
DR WPI: 1997-044837/05.	
DR N-PSDB; AAT73236.	
XX	
PT DNA of lactic acid bacteria - encoding enzymes involved in	
PT exo-polysaccharide biosynthesis	
XX	
PS Claim 7; Page 14-28; 46pp; French.	
XX	
CC This sequence represents the protein encoded by the epsI gene from the	
CC Streptococcus thermophilus exopolysaccharide (EPS) synthesis operon.	
CC The operon has 13 open reading frames (ORF) designated epsA-M encoding	
CC proteins AAW22165-81. This protein has 24% identity to a protein	
CC encoded by the RbV ORF of the rib cluster from Salmonella typhimurium	
CC and is probably a glycosyl transferase. The EPS are useful for	
CC thickening e.g. drinks, liquid desserts, yoghurts, soups, ice-creams,	
CC coffee creams, sauces, mayonnaise.	
XX	
SO Sequence 324 AA:	
Query Match	11.9%; Score 217.5; DB 18; Length 324;
Best Local Similarity	23.6%; Pred. No. 3-2e-14;
Matches 80; Conservative 67; Mismatches 103; Indels 89; Gaps 15;	

```

1 MQLPLSVLCAINVEKRYEQRASLAANVNOTWRNLILYDDGSDTGLTAIAKFOKRDRI 60
:::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1
3 LKSLISIAIPVNVKRYEKCLASVOAOTYNNFVILVNDGSDSLSTICEKFFVNDKRF 62

```


CC eps8, eps9 and eps10) that are involved in the biosynthesis of
CC exopolysaccharides (EPS). The enzymes catalyse the formation of
CC specific intersugar bonds. The enzymes catalyse a process that includes
CC at least one step of forming a bond (in alpha or beta anomeric form)
CC between C1, carrying the reducing aldehyde group of an activated D-Galp
CC (galactose in pyranose form), and a phosphate group on a lipophilic or
CC proteinaceous primer. The enzymes are used to produce EPS that have
CC antimicrobial or probiotic properties or are used in foodstuffs to improve
CC organoleptic qualities and flavour. When expressed by lactic acid
CC bacteria, EPS impart a free-flowing character and/or a smooth, creamy
CC texture to acidified milk products (yoghurt or cheese).

50 Sequence 316 AA;

Query Match	11.8%;	Score 214.5;	DB 21;	Length 316;
Best Local Similarity	35.7%;	Pred. No. 6.5e-14;		
Matches 45;	Conservative 29;	Mismatches 45;	Indels 7;	Gaps 2

```

Oy  1  MOPIVSVICIANENKEFEAGSLAVNOMQNRINDIILVDGSDRGCLAIKNDPOKRSRI 60
    1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1  MNPLISITVPEYNEKTIIRCTESILAQYTRNIEVILVNDGSDTQSLAVLSDLICSHNI 60
Oy  61  KIILAOANSGILPSLNGIDELAKSGGGGEYIARTADDIASPGWLEKIVGEMENDRST 120
    1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  61  KVIIMQ-KKQGSIVARNRTGID-----AATGKIYIAFVADDDKIRPEFVSSLYQTADKTGAD 113
Oy  121  TAMGAM 126
    1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  114  IVRGSF 119
    1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15
AAB47426
ID AAB47426 standard; Protein; 332 AA.

AC AAB47426;

DT 17-OCT-2001 (first entry)

DE EPSN

KW EPS plasmid; *Lactococcus lactis* subspecies *cremoris* Ropy352;
KW exopolysaccharide; EPS352; milk; thickener; glycosyltransferase;
KW biosynthesis; fruit juice.

05 Lactococcus lactis.

PN W0200157234-A2.

PD 09-AUG-2001.

PF 02-FEB-2001; 2001WO-US03404

PR 02-FEB-2000; 2000US-0179888.

PR 16-OCT-2000; 2000US-0241098.

PA (UYOR-) UNIV OREGON STATE.

PA (USDA) US DEPT OF AGRICULTURE.

PI Trempey JE, Knoshaug EP, Sandine WE, Ahlgren JA, Dierksen KP;

DR WPI; 2001-488889/53.

DR N-PSDB; AAH43198.

PT New bacterium useful in pharmaceutical formulations, food products and
PT beauty cosmetics, comprises characteristics of *Lactococcus lactis*
PT *cremoris* Ropy 352 -

PS Claim 28; Page 67-68; 73pp; English.

CC The sequences given in AAB47426-27 are encoded by a fragment of the EPS
CC plasmid derived from *L. lactis* subspecies *cremoris* Ropy352. The EPS
CC plasmid is about 32 kb in size and encodes at least 13 active genes.

The enzymes encoded by these genes allow the bacteria to produce an expopolysaccharide, designated EPS352. When EPS352 is expressed in or added to milk, it imparts highly desirable sensory characteristics to the milk, including making the milk very thick, with a very smooth mouth-feel, and slightly sweet with an obvious chewable-bite. Open reading frames (ORF's) M and N show homology to glycosyltransferase involved in EPS352 biosynthesis. *S. lactis cremoris* Ropy352 is deposited with the USDA-ARS-NCAR-NRRL as deposit accession number NRRL-B-302029. EPS352 is useful for thickening a liquid selected from CC milk, a milk-based liquid, a whey-based liquid, a soy-based liquid, CC and a fruit-juice. It is also useful as an additive in pharmaceutical products, beauty care products and coating agents.

SQ Sequence 332 AA;

Query Match	11.6%;	Score 211;	DB 22;	length 332;
Best Local Similarity	24.5%;	Pred. No. 1.6e-13;		
Matches 60;	Conservative 51;	Mismatches 84;	Indels 50;	Gaps 77;

```

QY      4 LVSIVLIVANEKYEFAOSLAIVNQTNRNIDILVDGSGDTGTLATAKDPQRKDSRIKIL 63
D      5 LITIIIVPYNEKTLRAIHSLNQTQNIIEVILINDSGIDSGEELISSQKKDKRK-L 63
QY      64 AQAONSGILPSLNLGDELAKSGGGGGGYIARTDADIASPGWITEKIVGEMEKDRSIAM 123
D      64 YNFKNLGVSHARNYGDIRE-----ASSGYITFELDPDDTYDKSCLEMGILNKFNADVVM 117
QY      124 GAW-----LE---VISEEKD-GKRLAKHHKKGKIMKPKPRHEDIAFFP 163
D      118 SNVYICCKGNIVPVVNDLLECEGLSLRDTMRSILSDTSFKGEVW----- 163
QY      164 FGNPIHNHTMIMRSRVIDGLRQYTERDEMADEYQFWVDYKLGRLAYPEALVKYRLIHAN 223
D      164 -----TRFERKNVNL-NVKFNEGISINYLEDMFLNFSIYNHNRILAYTKRKRIYFLQRE 214
QY      224 QVSSK 228
D      215 DSASK 219

```

Search completed: December 2, 2002, 11:58:59
Job time : 38.1203 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 : Search time 12.9535 Seconds
(without alignments)
790.458 Million cell updates/sec

Title: US-10-007-267-3

Perfect score: 1823

Sequence: 1 MGPLSVILICANNVEKYFAQ.....RLIKNRQAFSDSGKEGEI 348

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2.6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2.6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2.6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2.6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2.6/ptodata/1/1aa/PC10S.COMB.pep:*
6: /cgn2.6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1823	100.0	348	1	US-08-312-387B-3
2	1823	100.0	348	1	US-08-683-426-3
3	1823	100.0	348	1	US-08-683-458-3
4	1823	100.0	348	1	US-08-878-360-3
5	1823	100.0	348	4	US-09-333-412-3
6	1820	99.8	348	1	US-08-312-387B-11
7	1820	99.8	348	1	US-08-683-426-11
8	1820	99.8	348	1	US-08-683-458-11
9	1820	99.8	348	2	US-08-878-360-11
10	1820	99.8	348	3	US-08-478-140B-3
11	1820	99.8	348	4	US-08-478-140B-8
12	1820	99.8	348	4	US-09-333-412-11
13	1820	99.8	348	4	US-09-338-943-3
14	1820	99.8	348	4	US-09-338-943-8
15	1820	99.8	348	4	US-09-312-387B-5
16	1820	99.8	348	4	US-08-683-426-5
17	1820	99.8	348	4	US-08-683-458-5
18	1820	99.8	348	4	US-08-878-360-5
19	1820	99.8	348	4	US-09-333-412-5
20	1820	99.8	348	4	US-08-312-387B-12
21	1820	99.8	348	4	US-08-683-426-12
22	1820	99.8	348	4	US-08-683-458-12
23	1820	99.8	348	4	US-08-878-360-12
24	1820	99.8	348	4	US-08-478-140B-5
25	1820	99.8	348	4	US-09-333-412-12
26	1820	99.8	348	4	US-09-338-943-5
27	1820	99.8	348	4	US-08-597-236-10

28	217.5	11.9	324	1	US-08-746-682A-10	Sequence 10, Appl
29	210	11.5	965	4	US-09-437-277-3	Sequence 3, Appl
30	200	11.0	702	4	US-09-437-277-1	Sequence 1, Appl
31	191.5	10.5	674	4	US-08-961-083-200	Sequence 200, App
32	175	9.6	270	4	US-08-961-083-198	Sequence 198, App
33	175	9.6	281	4	US-08-961-083-196	Sequence 196, App
34	154.5	8.5	79	4	US-08-858-207A-521	Sequence 521, App
35	147.5	8.1	93	4	US-08-961-083-168	Sequence 168, App
36	130	7.1	358	4	US-09-134-001C-5633	Sequence 5633, Ap
37	129	7.1	303	4	US-08-961-083-202	Sequence 202, App
38	124.5	6.8	727	4	US-09-134-001C-4067	Sequence 4067, Ap
39	113.5	6.2	517	2	US-08-967-508-19	Sequence 19, Appl
40	113.5	6.2	517	2	US-08-967-506-19	Sequence 19, Appl
41	113.5	6.2	517	5	PCR-US94-02552-19	Sequence 19, Appl
42	113.5	6.2	559	2	US-08-967-508-9	Sequence 9, Appl
43	113.5	6.2	559	2	US-08-967-506-9	Sequence 9, Appl
44	113.5	6.2	559	5	PCR-US94-02552-9	Sequence 9, Appl
45	112.5	6.2	418	4	US-09-134-001C-4051	Sequence 4051, Ap

ALIGNMENTS

```
RESULT 1
US-08-312-387B-3
; Sequence 3, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauder & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-312-387B-3

Query Match 100.0%; Score 1823; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 5.5e-150;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLSVILICANNVEKYFAQSLAAVYNQWRNLDLIYVDSGTDLTAIAKPKQRDSRI 60
DB 1 MGPLSVILICANNVEKYFAQSLAAVYNQWRNLDLIYVDSGTDLTAIAKPKQRDSRI 60
QY 61 KILAAQNSGLIPISINIGHDELAKSGGGGEYIARTDADDIASPGWIERIVGEMEKDRSI 120
```

|||||
Db KILIAQONSLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDSRI 120
QY 121 IAMGAWLEVLSEKDGKGNRLARHHKGIWKKPTRHEDIAAFPPFGNPJHNTMIMRSYI 180
Db 121 IAMGAWLEVLSEKDGKGNRLARHHKGIWKKPTRHEDIAAFPPFGNPJHNTMIMRSYI 180
QY 181 DGLRLDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEITAOGI 240
Db 181 DGLRLDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEITAOGI 240
QY 241 OKTARNDFLOSGMFKTRFDSLEYROTAKAAYELPEKDLPEEDFEARARFLYOCFKRTDP 300
Db 241 OKTARNDFLOSGMFKTRFDSLEYROTAKAAYELPEKDLPEEDFEARARFLYOCFKRTDP 300
QY 301 PSGAWLDFADGRMRRLFTLROYFGILYRLIKNRROARSDSAGKEOEI 348
Db 301 PSGAWLDFADGRMRRLFTLROYFGILYRLIKNRROARSDSAGKEOEI 348

RESULT 2

US-08-683-426-3
; Sequence 3, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,426
; FILING DATE:
; CLASSIFICATION: 536
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-426-3

Query Match 100.0%; Score 1823; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 5,5e-190;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPIVSVLICAVNVEKYFAQSILAAVYVNTWRNLDILIVDGSITDGLTALAKDFOKRDSRI 60
Db 1 MOPIVSVLICAVNVEKYFAQSILAAVYVNTWRNLDILIVDGSITDGLTALAKDFOKRDSRI 60

QY 61 KILIAQONSLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDSRI 120
Db 61 KILIAQONSLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDSRI 120
QY 121 IAMGAWLEVLSEKDGKGNRLARHHKGIWKKPTRHEDIAAFPPFGNPJHNTMIMRSYI 180
Db 121 IAMGAWLEVLSEKDGKGNRLARHHKGIWKKPTRHEDIAAFPPFGNPJHNTMIMRSYI 180
QY 181 DGLRLDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEITAOGI 240
Db 181 DGLRLDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEITAOGI 240
QY 241 OKTARNDFLOSGMFKTRFDSLEYROTAKAAYELPEKDLPEEDFEARARFLYOCFKRTDP 300
Db 241 OKTARNDFLOSGMFKTRFDSLEYROTAKAAYELPEKDLPEEDFEARARFLYOCFKRTDP 300
QY 301 PSGAWLDFADGRMRRLFTLROYFGILYRLIKNRROARSDSAGKEOEI 348
Db 301 PSGAWLDFADGRMRRLFTLROYFGILYRLIKNRROARSDSAGKEOEI 348

RESULT 3

US-08-683-458-3
; Sequence 3, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-458-3

Query Match 100.0%; Score 1823; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 5,5e-190;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPIVSVLICAVNVEKYFAQSILAAVYVNTWRNLDILIVDGSITDGLTALAKDFOKRDSRI 60
Db 1 MOPIVSVLICAVNVEKYFAQSILAAVYVNTWRNLDILIVDGSITDGLTALAKDFOKRDSRI 60

QY	61	KILQAOANUSGLIPSLNIGLDELASGGGGGYTARTADDIASGWTKEKIVGEMKDRSI	120
QY	61	KILQAOANUSGLIPSLNIGLDELASGGGGGYTARTADDIASGWTKEKIVGEMKDRSI	120
Db	61	KILQAOANUSGLIPSLNIGLDELASGGGGGYTARTADDIASGWTKEKIVGEMKDRSI	120
QY	121	IAMGAMLEVLISEEKDGNRLAHNRHKGKIMKPTTHEDIAAFPEFGNFIHNNTIMRASYI	180
Db	121	IAMGAMLEVLISEEKDGNRLAHNRHKGKIMKPTTHEDIAAFPEFGNFIHNNTIMRASYI	180
QY	181	DGGLRYDTERBWMADYQFWVDVSKSLGSLAYYPRLAYYRLAHNAVSSKHSVRHETAOGI	240
Db	181	DGGLRYDTERBWMADYQFWVDVSKSLGSLAYYPRLAYYRLAHNAVSSKHSVRHETAOGI	240
QY	241	OKTARNDFLQSMGKTRFDSLEYRQTKAAAYELPEKDLPEDEFERARFLEYQCFKRTDTP	300
Db	241	OKTARNDFLQSMGKTRFDSLEYRQTKAAAYELPEKDLPEDEFERARFLEYQCFKRTDTP	300
QY	301	PSGAMLDPAADGNRRRLFTLYROYGILYLRLKKNRQARSACKEDPT	348
Db	301	PSGAMLDPAADGNRRRLFTLYROYGILYLRLKKNRQARSACKEDPT	348

RESULT 4

```

US-08-878-360-3
Sequence 3, Application US/08878360
Patent No. 5945322
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/683,426
FILING DATE:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 467-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-3

```

Query Match	100.0%	Score 1823	DB 2	Length 348
Best Local Similarity	100.0%	Pred. No. 1,55e-190		
Matches 348; Conservative	0;	Mismatches	0;	Gaps 0

QY	1	MOPLYSVICAYNEKEFAOGLAAVVMOTNRNDILLVDDSGTGTALIKDKOQRORSRI	60
Dp	1	MOPLYSVICAYNEKEFAOGLAAVVMOTNRNDILLVDDSGTGTALIKDKOQRORSRI	60
QY	61	KILAQONSGILPSLNTIGLDELAKSGGGGGEYIARTADDIASPGWIEKTVGEMEKDRSI	120
Dp	61	KILAQONSGILPSLNTIGLDELAKSGGGGGEYIARTADDIASPGWIEKTVGEMEKDRSI	120
QY	121	IAMGWLVELEEEEDGNRLAHNRHKGKIMKPPTRHEDIAAFPPGPNYIHNTMTMRRSVI	180
Dp	121	IAMGWLVELEEEEDGNRLAHNRHKGKIMKPPTRHEDIAAFPPGPNYIHNTMTMRRSVI	180
QY	181	DGGLRTDERDWAEDYQFWYDVSKLGRLAYIPEALVYKRLHANOVSSKHSVROHEIQAOTI	240
Dp	181	DGGLRTDERDWAEDYQFWYDVSKLGRLAYIPEALVYKRLHANOVSSKHSVROHEIQAOTI	240
QY	241	OKTARNDFLQSMGEKFTFSDSLEYKOTKAAAYELPEKPLPEEDFEFARARFLYOCFFKRTDTP	300
Dp	241	OKTARNDFLQSMGEKFTFSDSLEYKOTKAAAYELPEKPLPEEDFEFARARFLYOCFFKRTDTP	300
QY	301	PSGAWLDFADGRRMRRLFTLYQYIGILYRLIKNRQARSACKQEOTI	348
Dp	301	PSGAWLDFADGRRMRRLFTLYQYIGILYRLIKNRQARSACKQEOTI	348

RESULT 5

US-09-333-412-3
Sequence 3, Application US/09333412
Patent No. 6342382
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYN-
OLIGOSACCHARIDES, AND GENES ENCO-
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-333-412-3

Query Match	100.0%;	Score 1823;	DB 4;	Length 348;
Best Local Similarity	100.0%;	Pred. No. 5.5e-190;		
Matches 348;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Category	Conservative	Mismatches	Indels	Gaps
1 MODIFIER TO VARIATION	34/1	0/1	0/1	0/1


```

:|||||
Db 1 LQPLVSLICAVNVEKYFAQSLAAVYNQWTNRNLDLIVDGSITDGTALAKDFQKRSRI 60
Qy 61 KILAAQNSGLIPSLINIGIDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
Db 61 KILAAQNSGLIPSLINIGIDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
Qy 121 IIMGAWLEVLSEKDGNRILARRHHKGIWKKPTRHEDIAAFPEFGNPJNNNTMIMRSYI 180
Db 121 IIMGAWLEVLSEKDGNRILARRHHKGIWKKPTRHEDIAAFPEFGNPJNNNTMIMRSYI 180
Qy 181 DGLRYDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANQVSSKHSVROHEIAQGI 240
Db 181 DGLRYDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANQVSSKHSVROHEIAQGI 240
Qy 241 OKTARNDFLOSNGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDTP 300
Db 241 OKTARNDFLOSNGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDTP 300
Qy 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEQEI 348
Db 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEQEI 348

```

RESULT 8

```

US-08-683-458-11
; Sequence 11, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-683-458-11

```

Query Match 99.8%; Score 1820; DB 1; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 MOPVSLVICAVNVEKYFAQSLAAVYNQWTNRNLDLIVDGSITDGTALAKDFQKRSRI 60
:|||||
Db 1 LQPLVSLICAVNVEKYFAQSLAAVYNQWTNRNLDLIVDGSITDGTALAKDFQKRSRI 60
Qy 61 KILAAQNSGLIPSLINIGIDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
Db 61 KILAAQNSGLIPSLINIGIDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
Qy 121 IIMGAWLEVLSEKDGNRILARRHHKGIWKKPTRHEDIAAFPEFGNPJNNNTMIMRSYI 180
Db 121 IIMGAWLEVLSEKDGNRILARRHHKGIWKKPTRHEDIAAFPEFGNPJNNNTMIMRSYI 180
Qy 181 DGLRYDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANQVSSKHSVROHEIAQGI 240
Db 181 DGLRYDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANQVSSKHSVROHEIAQGI 240
Qy 241 OKTARNDFLOSNGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDTP 300
Db 241 OKTARNDFLOSNGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDTP 300
Qy 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEQEI 348
Db 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEQEI 348

```

RESULT 9

```

US-08-878-360-11
; Sequence 11, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,360
; FILING DATE: 18-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,426
; FILING DATE:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-878-360-11

```

Query Match 99.8%; Score 1820; DB 2; Length 348;

Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLICAVNEKYEKQSLAAVYVQWTRNDILIVDGGSTDTGLATAKPFQKRDSTI 60
Db 1 LOPVSVLICAVNEKYEKQSLAAVYVQWTRNDILIVDGGSTDTGLATAKPFQKRDSTI 60
QY 61 KILAAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
Db 61 KILAAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
QY 121 IAMGAWLEVLSEKDGKRLARHHKGIKKKPTRHEDIAAFPPGPNJHNTMIMRSYI 180
Db 121 IAMGAWLEVLSEKDGKRLARHHKGIKKKPTRHEDIAAFPPGPNJHNTMIMRSYI 180
QY 181 DGLRDTEDMAEDYQFYDVYSKGLRLAYYPALVKYRLHANOVSSKHSVRQHEIAOGI 240
Db 181 DGLRDTEDMAEDYQFYDVYSKGLRLAYYPALVKYRLHANOVSSKHSVRQHEIAOGI 240
QY 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDPERARRFLYOCFKRTDP 300
Db 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDPERARRFLYOCFKRTDP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
Db 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348

RESULT 10

US-08-478-140B-3
; Sequence 3, Application US/08478140B
; Patent No. 6127153

GENERAL INFORMATION:

APPLICANT: JOHNSON, KARL F.
APPLICANT: BOTH, STEPHEN
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,140B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-140B-3

Query Match 99.8%; Score 1820; DB 3; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLICAVNEKYEKQSLAAVYVQWTRNDILIVDGGSTDTGLATAKPFQKRDSTI 60
Db 1 LOPVSVLICAVNEKYEKQSLAAVYVQWTRNDILIVDGGSTDTGLATAKPFQKRDSTI 60
QY 61 KILAAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
Db 61 KILAAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
QY 121 IAMGAWLEVLSEKDGKRLARHHKGIKKKPTRHEDIAAFPPGPNJHNTMIMRSYI 180
Db 121 IAMGAWLEVLSEKDGKRLARHHKGIKKKPTRHEDIAAFPPGPNJHNTMIMRSYI 180
QY 181 DGLRDTEDMAEDYQFYDVYSKGLRLAYYPALVKYRLHANOVSSKHSVRQHEIAOGI 240
Db 181 DGLRDTEDMAEDYQFYDVYSKGLRLAYYPALVKYRLHANOVSSKHSVRQHEIAOGI 240
QY 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDPERARRFLYOCFKRTDP 300
Db 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDPERARRFLYOCFKRTDP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
Db 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348

RESULT 11

US-08-478-140B-8
; Sequence 8, Application US/08478140B
; Patent No. 6127153

GENERAL INFORMATION:

APPLICANT: JOHNSON, KARL F.
APPLICANT: BOTH, STEPHEN
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,140B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein

ANTI-SENSE: NO
US-08-478-140B-8

Query Match 99.8%; Score 1820; DB 3; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLICAVNVEKFFQSLAAVYVNTWRNLDLIYDGSSTDTGLTAKDFOKRDSRI 60
DB 1 LOPVSVLICAVNVEKFFQSLAAVYVNTWRNLDLIYDGSSTDTGLTAKDFOKRDSRI 60
QY 61 KILAAQNSGLIPSLNIGDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
DB 61 KILAAQNSGLIPSLNIGDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
QY 121 IAMGAMLEVLSEKDGKGNRLARHHKHKIWKKPTRHEDIAAFPPGPNHNMTMIMRSYI 180
DB 121 IAMGAMLEVLSEKDGKGNRLARHHKHKIWKKPTRHEDIAAFPPGPNHNMTMIMRSYI 180
QY 181 DGLRDTFRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQSI 240
DB 181 DGLRDTFRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQSI 240
QY 241 OKTARNDFLQSMGFRTFDSLEYROTAKAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
DB 241 OKTARNDFLQSMGFRTFDSLEYROTAKAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKEDEI 348
DB 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKEDEI 348

RESULT 12
US-09-333-412-11

Sequence 11, Application US/09333412
Patent No. 6342382

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412

FILING DATE: 15-Jun-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-095

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 348 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-333-412-11

Query Match 99.8%; Score 1820; DB 4; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLICAVNVEKFFQSLAAVYVNTWRNLDLIYDGSSTDTGLTAKDFOKRDSRI 60
DB 1 LOPVSVLICAVNVEKFFQSLAAVYVNTWRNLDLIYDGSSTDTGLTAKDFOKRDSRI 60
QY 61 KILAAQNSGLIPSLNIGDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
DB 61 KILAAQNSGLIPSLNIGDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
QY 121 IAMGAMLEVLSEKDGKGNRLARHHKHKIWKKPTRHEDIAAFPPGPNHNMTMIMRSYI 180
DB 121 IAMGAMLEVLSEKDGKGNRLARHHKHKIWKKPTRHEDIAAFPPGPNHNMTMIMRSYI 180
QY 181 DGLRDTFRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQSI 240
DB 181 DGLRDTFRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQSI 240
QY 241 OKTARNDFLQSMGFRTFDSLEYROTAKAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
DB 241 OKTARNDFLQSMGFRTFDSLEYROTAKAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKEDEI 348
DB 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKEDEI 348

RESULT 13
US-09-338-943-3

Sequence 3, Application US/09338943

Patent No. 6379933

GENERAL INFORMATION:

APPLICANT: JOHNSON, KARL F.

APPLICANT: ROTH, STEPHEN

APPLICANT: BUCZALA, STEPHANIE L.

TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO

TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A

TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A

TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/338,943

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/478,140

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Laura A. Coruzzi

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7188-017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

;; INFORMATION FOR SEQ ID NO: 3 :
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 348 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-338-943-3

Query Match 99.8%; Score 1820; DB 4; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLVSVLICAVNVEKFFAQSIAAVVNOTWRMDLIYDDGSTDGLTAAKDFOKRDSRI 60
:|||||
DB 1 LQPLVSVLICAVNVEKFFAQSIAAVVNOTWRMDLIYDDGSTDGLTAAKDFOKRDSRI 60
QY 61 KILAAQONSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
:|||||
DB 61 KILAAQONSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
QY 121 IAMGAWLEVLSEKQGNRLARHHKHKIWKKPTRHEDIAAFPPGPNPHNNTMIMRSYI 180
:|||||
DB 121 IAMGAWLEVLSEKQGNRLARHHKHKIWKKPTRHEDIAAFPPGPNPHNNTMIMRSYI 180
QY 181 DGLGRVDFERDMAEDYQFYDYVSKLGRLAYYPEALVKYRLHANOVSSKHSVRQHEIAQSI 240
:|||||
DB 181 DGLGRVDFERDMAEDYQFYDYVSKLGRLAYYPEALVKYRLHANOVSSKHSVRQHEIAQSI 240
QY 241 OKTARNDFLOSMGFTRFDSLEYRQTKAAAYELPEKDLPEEDFEERARFLYOCFFKRTDP 300
:|||||
DB 241 OKTARNDFLOSMGFTRFDSLEYRQTKAAAYELPEKDLPEEDFEERARFLYOCFFKRTDP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKDEOI 348
:|||||
DB 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKDEOI 348

RESULT 14
US-09-338-943-8
; Sequence 8, Application US/09338943
; Patent No. 6379933
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: ROTH, STEPHEN
; APPLICANT: BOCCAZZA, STEPHANIE L.
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/338,943
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/478,140
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-9741/8864
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 8 :
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 348 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
; AMT-SENSE: NO
US-09-338-943-8

Query Match 99.8%; Score 1820; DB 4; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLVSVLICAVNVEKFFAQSIAAVVNOTWRMDLIYDDGSTDGLTAAKDFOKRDSRI 60
:|||||
DB 1 LQPLVSVLICAVNVEKFFAQSIAAVVNOTWRMDLIYDDGSTDGLTAAKDFOKRDSRI 60
QY 61 KILAAQONSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
:|||||
DB 61 KILAAQONSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
QY 121 IAMGAWLEVLSEKQGNRLARHHKHKIWKKPTRHEDIAAFPPGPNPHNNTMIMRSYI 180
:|||||
DB 121 IAMGAWLEVLSEKQGNRLARHHKHKIWKKPTRHEDIAAFPPGPNPHNNTMIMRSYI 180
QY 181 DGLGRVDFERDMAEDYQFYDYVSKLGRLAYYPEALVKYRLHANOVSSKHSVRQHEIAQSI 240
:|||||
DB 181 DGLGRVDFERDMAEDYQFYDYVSKLGRLAYYPEALVKYRLHANOVSSKHSVRQHEIAQSI 240
QY 241 OKTARNDFLOSMGFTRFDSLEYRQTKAAAYELPEKDLPEEDFEERARFLYOCFFKRTDP 300
:|||||
DB 241 OKTARNDFLOSMGFTRFDSLEYRQTKAAAYELPEKDLPEEDFEERARFLYOCFFKRTDP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKDEOI 348
:|||||
DB 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKDEOI 348

RESULT 15
US-08-312-387B-5
; Sequence 5, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 12:00:15 : Search time 6.97495 Seconds
(without alignments)
794.504 Million cell updates/sec

Title: US-10-007-267-3
Perfect score: 1823
Sequence: 1 MQLPVSYLICAYNTEKYFAQ.....RLIKNRQARSASAKKEGEI 348

Scoring table:
BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCr_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCrUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1823	100.0	348	12	US-10-007-267-3
2	1820	99.8	348	12	US-10-007-267-11
3	1049	57.5	337	12	US-10-007-267-5
4	1046	57.4	337	12	US-10-007-267-12
5	230	12.6	150	10	US-09-924-358-29
6	228.5	12.5	332	10	US-09-767-041-22
7	227	12.5	278	10	US-09-767-041-36
8	210.5	11.5	313	10	US-09-900-038A-1
9	210	11.5	322	10	US-09-767-041-34
10	208.5	11.4	972	9	US-09-879-959-10
11	195	10.7	301	10	US-09-816-028A-27
12	194.5	10.7	322	10	US-09-767-041-35
13	194	10.6	303	10	US-09-816-028A-29
14	191.5	10.5	674	10	US-09-765-272-200
15	190	10.4	270	10	US-09-816-028A-39
16	183	10.0	332	10	US-09-767-041-21
17	179	9.8	389	10	US-09-816-028A-34
18	176	9.7	120	10	US-09-767-041-52
19	175	9.6	270	10	US-09-765-272-198

20	175	9.6	281	10	US-09-765-272-196	Sequence 196, App
21	169	9.3	297	10	US-09-816-028A-31	Sequence 31, Appl
22	168.5	9.2	120	10	US-09-767-041-51	Sequence 51, Appl
23	164	9.0	706	10	US-09-815-242-4950	Sequence 4950, Ap
24	164	9.0	715	10	US-09-815-242-10511	Sequence 10511, A
25	158	8.7	210	10	US-09-767-041-47	Sequence 47, Appl
26	151	8.3	187	9	US-09-973-457-4	Sequence 4, Appl
27	151	8.3	187	10	US-09-815-028-7	Sequence 7, Appl
28	151	8.3	187	12	US-10-074-527-4	Sequence 4, Appl
29	147.5	8.1	79	10	US-09-765-272-168	Sequence 168, App
30	142.5	7.8	269	10	US-09-767-041-41	Sequence 41, Appl
31	129	7.1	303	10	US-09-765-272-202	Sequence 202, App
32	123	6.7	358	10	US-09-815-242-5714	Sequence 5714, Ap
33	123	6.7	573	10	US-09-815-242-12474	Sequence 12474, A
34	116.5	6.4	559	12	US-10-001-851-22	Sequence 22, Appl
35	113.5	6.2	559	12	US-10-001-851-20	Sequence 20, Appl
36	113.5	6.2	559	12	US-10-001-851-21	Sequence 21, Appl
37	113.5	6.2	559	12	US-10-001-851-23	Sequence 23, Appl
38	113.5	6.2	561	10	US-09-925-301-1006	Sequence 1006, Ap
39	111.5	6.1	612	12	US-10-001-851-25	Sequence 25, Appl
40	110.5	6.1	559	12	US-10-001-851-24	Sequence 24, Appl
41	104	5.7	418	10	US-09-816-028A-33	Sequence 33, Appl
42	102	5.6	626	12	US-10-001-851-27	Sequence 27, Appl
43	100.5	5.5	608	10	US-09-924-358-8	Sequence 8, Appl
44	98.5	5.4	256	10	US-09-925-301-883	Sequence 883, App
45	97.5	5.3	326	9	US-10-108-605-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-10-007-267-3
Sequence 3, Application US/10007267
Patent No. US20020127682A1
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-007-267-3

Query Match 100.0%; Score 1823; DB 12; Length 348;
Best Local Similarity 100.0%; Pred. No. 1.9e-163;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLVSVLICAVVEKFFQSLAAVYNQTMRLDILIVDGSIDGTLATAKDFQKDSRI 60
DB 1 MQLVSVLICAVVEKFFQSLAAVYNQTMRLDILIVDGSIDGTLATAKDFQKDSRI 60
QY 61 KILAQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGMIETKIVGEMEKDRSI 120
DB 61 KILAQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGMIETKIVGEMEKDRSI 120
QY 121 IANGAMLEVLSEKDGKRNRLARHHKHKIKMKPTRHEDIAAFPPFGNPINHTMIMRSVI 180
DB 121 IANGAMLEVLSEKDGKRNRLARHHKHKIKMKPTRHEDIAAFPPFGNPINHTMIMRSVI 180
QY 181 DGLRVDTERDMADYQFWYDVSKLGLRLAYYPEALVKYRLHANOVSSKSHVRQHEIAGCI 240
DB 181 DGLRVDTERDMADYQFWYDVSKLGLRLAYYPEALVKYRLHANOVSSKSHVRQHEIAGCI 240
QY 241 OKTARNDFLOSMGFKTRFDSLERYRQTKAAAYELPEKDLPEDEFEARRRFLYOCFKRTDTP 300
DB 241 OKTARNDFLOSMGFKTRFDSLERYRQTKAAAYELPEKDLPEDEFEARRRFLYOCFKRTDTP 300
QY 301 PSGAWLDFPADGGRMRRLFTLRQYFGILYRLIKNRROARSDSAGKEOEI 348
DB 301 PSGAWLDFPADGGRMRRLFTLRQYFGILYRLIKNRROARSDSAGKEOEI 348

RESULT 2

US-10-007-267-11
Sequence 11, Application US/10007267
Patent No. US20020127682A1

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF OLIGOSACCHARIDES, AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,267

FILING DATE: 03-Dec-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412

FILING DATE: 15-Jun-1999

APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-095

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-007-267-11

Query Match 99.8%; Score 1820; DB 12; Length 348;
Best Local Similarity 99.7%; Pred. No. 3.7e-163;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLVSVLICAVVEKFFQSLAAVYNQTMRLDILIVDGSIDGTLATAKDFQKDSRI 60
DB 1 MQLVSVLICAVVEKFFQSLAAVYNQTMRLDILIVDGSIDGTLATAKDFQKDSRI 60
QY 61 KILAQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGMIETKIVGEMEKDRSI 120
DB 61 KILAQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGMIETKIVGEMEKDRSI 120
QY 121 IANGAMLEVLSEKDGKRNRLARHHKHKIKMKPTRHEDIAAFPPFGNPINHTMIMRSVI 180
DB 121 IANGAMLEVLSEKDGKRNRLARHHKHKIKMKPTRHEDIAAFPPFGNPINHTMIMRSVI 180
QY 181 DGLRVDTERDMADYQFWYDVSKLGLRLAYYPEALVKYRLHANOVSSKSHVRQHEIAGCI 240
DB 181 DGLRVDTERDMADYQFWYDVSKLGLRLAYYPEALVKYRLHANOVSSKSHVRQHEIAGCI 240
QY 241 OKTARNDFLOSMGFKTRFDSLERYRQTKAAAYELPEKDLPEDEFEARRRFLYOCFKRTDTP 300
DB 241 OKTARNDFLOSMGFKTRFDSLERYRQTKAAAYELPEKDLPEDEFEARRRFLYOCFKRTDTP 300
QY 301 PSGAWLDFPADGGRMRRLFTLRQYFGILYRLIKNRROARSDSAGKEOEI 348
DB 301 PSGAWLDFPADGGRMRRLFTLRQYFGILYRLIKNRROARSDSAGKEOEI 348

RESULT 3

US-10-007-267-5
Sequence 5, Application US/10007267
Patent No. US20020127682A1

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF OLIGOSACCHARIDES, AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,267

FILING DATE: 03-Dec-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412

FILING DATE: 15-Jun-1999

APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-095

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-007-267-5

Query Match 57.5%; Score 1049; DB 12; Length 337;
Best Local Similarity 64.5%; Pred. No. 7.1e-91;
Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

QY 1 MOPVSVLICAVNVEKYFAQSIAAVVQNTWRNLDILYDDGSTDGTALAKDFQRDSRI 60
DB 1 MOPVSVLICAVNAEKYFAQSIAAVVQNTWRNLDILYDDGSTDGTALAKDFQRDSRI 60
QY 61 KILAAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDDIAPGMIKIVGEMEKDRSI 120
DB 61 RLISNPNRLGFASINIGDELAKS--GGGEYIARTDADDDIAPGMIKIVGEMEKDRSI 118
QY 121 IAMGAMLEVLSEKQGNRLARHHKHKTKPTRHEDIAAFPPGPNPHNNTMIMRSYI 180
DB 119 IAMGAMLEVLSEKQGNRLARHHKHKTKPTRHEDIAAFPPGPNPHNNTMIMRSYI 178
QY 181 DGLGLYTERDWAEDYQFQVSVKGLRLAYYPEALVKYRLHANOVSSKSHVROHEIAGI 240
DB 179 DGLGLYTERDWAEDYQFQVSVKGLRLAYYPEALVKYRLHANOVSSKSHVROHEIAGI 238
QY 241 OKTARNDELQSGFKTRPSLEYRQTKAAAYELPEKDLPEEDFERARRLYOCFRKTDP 300
DB 239 KEIRAGYWKAGIAGVADCLNYGLKSTAYALYERKALSGDIGCLRFLYFLYLEKY 298
QY 301 PSGAMLDFAADGRMRFLTRQYFGLYRLK 332
DB 299 SLTDLLDPLTDPRVMRKLFAPQYRKILKKMLR 330

RESULT 4
US-10-007-267-12

Sequence 12, Application US/10007267
Patent No. US20020127682A1
GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999

APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-007-267-12

Query Match 57.4%; Score 1046; DB 12; Length 337;
Best Local Similarity 64.2%; Pred. No. 1.4e-90;
Matches 213; Conservative 34; Mismatches 83; Indels 2; Gaps 1;

QY 1 MOPVSVLICAVNVEKYFAQSIAAVVQNTWRNLDILYDDGSTDGTALAKDFQRDSRI 60
DB 1 MOPVSVLICAVNAEKYFAQSIAAVVQNTWRNLDILYDDGSTDGTALAKDFQRDSRI 60
QY 61 KILAAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDDIAPGMIKIVGEMEKDRSI 120
DB 61 RLISNPNRLGFASINIGDELAKS--GGGEYIARTDADDDIAPGMIKIVGEMEKDRSI 118
QY 121 IAMGAMLEVLSEKQGNRLARHHKHKTKPTRHEDIAAFPPGPNPHNNTMIMRSYI 180
DB 119 IAMGAMLEVLSEKQGNRLARHHKHKTKPTRHEDIAAFPPGPNPHNNTMIMRSYI 178
QY 181 DGLGLYTERDWAEDYQFQVSVKGLRLAYYPEALVKYRLHANOVSSKSHVROHEIAGI 240
DB 179 DGLGLYTERDWAEDYQFQVSVKGLRLAYYPEALVKYRLHANOVSSKSHVROHEIAGI 238
QY 241 OKTARNDELQSGFKTRPSLEYRQTKAAAYELPEKDLPEEDFERARRLYOCFRKTDP 300
DB 239 KEIRAGYWKAGIAGVADCLNYGLKSTAYALYERKALSGDIGCLRFLYFLYLEKY 298
QY 301 PSGAMLDFAADGRMRFLTRQYFGLYRLK 332
DB 299 SLTDLLDPLTDPRVMRKLFAPQYRKILKKMLR 330

RESULT 5
US-09-924-358-29

Sequence 29, Application US/09924358
Patent No. US20020107376A1
GENERAL INFORMATION:

APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 26199, 33530, 33949, 47148, 50226, AND
TITLE OF INVENTION: 58764,
FILE REFERENCE: 38155-20034.00
CURRENT APPLICATION NUMBER: US/09/924,358
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/229,300
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29

LENGTH: 150

TYPE: prt
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Consensus amino acid
US-09-924-358-29

Query Match 12.6%; Score 230; DB 10; Length 150;
Best Local Similarity 40.8%; Pred. No. 1.5e-14;
Matches 53; Conservative 29; Mismatches 38; Indels 10; Gaps 5;

Query Match	12.5%	Score 228.5	DB 10	Length 332
Best Local Similarity	23.6%	Pred. No. 6.1e-14		
Matches	87	Conservative	68	Mismatches 99
			Indels 115	Gaps 16
OY	5	VSLICAYNVEKFAQSLAAYVNTQTRNDILIDVDGSTDGTLAIKDKQKRDSRIKILA	64	
	1			
	4	ISIIPIYNVEELKSCINSIYNQTYKHIEILLVNDGSDNSEECILAYAKKDSRIYF-	62	
OY	65	QAQNSGLPISLNIGIDELAKSGGGGGYIARPDADDIASPGMIEIKTVGEMEKDRIIANG	124	
	63	KKEKGLSDARNYGISR-AK-----GDYLAFLDSDFEIIHFSEIQRLEHAEIERENALVAVA	116	
OY	125	AMLEYLESEKDGKNRLARHHKHGKIMKKPRHEDIAAFPEFGAPRINHNTIMR-RSVIDGG	183	
	117	GIDRY-----DASHGFLTAERPLPNOAVLSGRNCKKL	149	
OY	184	LRDYTER--DW-----AADYOFWYD-VSKLGRLAYYPEALVYKRLH	221	
	150	LEADGHRFVAANKLYKKELDFREFEKGKIHDEYEPPTYRLYLELEKVALVKECLYYVDR	209	
OY	222	ANQV--SKSHVSRQHEIAGIOKTRANDPLQSGMGFTTRDSLEYRQTKAAAYELPKDLP	279	
	210	ENSIIITSSMTDHRFCLLE--FQNERMDYESRG-----DKEL-	245	
OY	280	EEDFERARFLIQQKRTDPPSGAMLDGAA-----DGRMRFLTLROYFGILVR	329	
	246	-----LLECIR-----SFLAAVAVLELCKYINWLSKQCKKLTLL--FRIYVK	284	
OY	330	LTK-NRROA	337	
Db	285	OLQKMKRLA	293	

Query Match	Similarity	12.5%	Score 227	DB 10	length 278
Best Local	Similarity 24.6%		Pred. No. 6.6e-14		
Matches	68	Conservative	60	Mismatches	88
				Indels	60
				Gaps	9
OY	5	VSLLICAVNVEKYFAQSLAAVAVVQTRNLDILVDDGSTDGTLLAKDFOKRDSRIKILA	64		
Db	7	ISLIPIYVNEVYKLSKCIDSIYNQTYKHIEILLVNDSDNSEELICLAAKDSRIIRYF	65		
OY	65	QANQSGILPSSLIGDELAKSGGGGEVYIARTDADDIAPGNIETKIVGEMKDRSIIANG	124		
Db	66	KKENGGLSPARKYIGISR-AK-----GDYIAFLDSDFIHSERIQRLHEIEENALVAVNA	119		
OY	125	AWLEYLSEEEKDGNRLARHHKHGKIMKPTRHEDIAAFPEFGNPJHNNMTIMR-RSYIDGG	183		
Db	120	GQDHY-----DASGHFLTAELPLPNOAVLGRNNCKRL	152		
OY	184	IKYDER-----DMAEDYOF-----WYDYSKGLRAYBEALKYRL	220		
Db	153	LEADGHRFVACNKLKYLKFLFEDFRPREKKIHDEDFYTRLLRYELEKVALIVECLYYVD	212		
OY	221	HANQVSKHSVROHEITAOIQ-KTARNDFLQSMGFK	255		
Db	213	RENSITTS-SMTDHFHCHLLEFQNEIMQFESRGDK	247		

```

RESULT 8
US-09-900-038A-1
; Sequence 1, Application US/09900038A
; Patent No. US20020142425A1
; GENERAL INFORMATION:
; APPLICANT: Miyake, Katsuhide
; APPLICANT: Watanabe, Masaki
; APPLICANT: Iijima, Shinji
; TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
; FILE REFERENCE: 766.53
; CURRENT APPLICATION NUMBER: US/09/900.038A
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: JP 2001-392
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae Type Ib
US-09-900-038A-1

```


Db 169 IYKAFNMN--LKKBIKINMADALLYPLTISNEIFYLQPLYTOHNSNSTNNIN 226
QY 226 SSKHSHVROHIAGIOKTARN 246
Db 227 SLEANTOEKIKVNLKSTKN 247

RESULT 14
US-09-765-272-200

; Sequence 200, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8504
; INFORMATION FOR SEQ ID NO: 200:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 674 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 200:
US-09-765-272-200

Query Match 10.5%; Score 191.5; DB 10; Length 674;
Best Local Similarity 21.2%; Pred. No. 4.7e-10;
Matches 68; Conservative 64; Mismatches 102; Indels 87; Gaps 13;

QY 22 LAAVVOTMNRNLDLIVDGSSTGTLAIKDPKRSRIKILAAQNSGLIPSLNIGLDE 81
Db 2 LDSIIITQTKNIIIVVNDSDTASGEICFSEMDHRIIYIQ-ENAGLSAARNGLNN 60
QY 82 LAKSGGGGEYIARTADIASPGWIEKIVGEMEKDRSIITAMGAMLEVLSEEDGNRLAR 141
Db 61 MS-----GNVVFVDSDDMIEGDYVELTKKIVEXQADIAGVNGYSF--NESEG--MEY 110
QY 142 HHKHGKIMKPTTHEDIAAFPPGPNIHNTMTIMRSVIDGILRYDTER-----DMA 193
Db 111 FHILGD-----SYTERKYDVMVSIFEN-----LYETQEKMSFALISAMG 148
QY 194 EDY-----QFWYDVSGLG-----RLAYYPEALVKYRLRHANQVSSKHSVR- 232
Db 149 KLYKALFEDQLRDIGKLGEDGILNOKVYILSKVYILNKSLEYAIRIRKGSLSRWTEKW 208
QY 233 OHEIAOGIOKTARNDFLOSNGMFKTRFDSLEYRO-----TKAAAYELPEKDL 278

Db 209 MHALVAMSE--RTITLLANGYPLEKHIAVYROMLEVSLANQASGLSPATATY----- 259
QY 279 PEEDFERARRELYQCFKRTDT 299
Db 260 --KEFEKORLLNQLSROES 278

RESULT 15
US-09-816-028A-39

; Sequence 39, Application US/09816028A
; Patent No. US20020042369A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter glycosyltransferases for Biosynthesis of
; FILE REFERENCE: 019633-00011105
; CURRENT APPLICATION NUMBER: US/09/816,028A
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: glycosyltransferase from C. jejuni OH4384 (ORF 12a
; US-09-816-028A-39

Query Match 10.4%; Score 190; DB 10; Length 270;
Best Local Similarity 22.0%; Pred. No. 1.9e-10;
Matches 66; Conservative 54; Mismatches 102; Indels 78; Gaps 9;

QY 3 PLYSLICATVNEKRYFRQSLAAVAVNOTWRNLDLIVDGSSTGTLAIKDPKRSRIKI 62
Db 2 POLSIIIPFNSCDFISRALQSCINOTLKDIETILIIDKSKDMSLNVLEFAKKDPRIKI 61
QY 63 LAQANSGLIPSLNIGLDELAKSGGGGEYIARTADIASPGWIEKIVGEMEKDRSIIA 122
Db 62 FQNEENLGTFFASRNGLVLR-----SSDFIMFLDSDDFLTPDACEIAFEKMKGFDLIC 115
QY 123 MGA-----WLEVLSEEDKGNRLARHHKHKGIWKKPTRHEDIA 159
Db 116 FDAFVHRVKTQFYRRKQDEVFNQKELEFLSKOR-----HFCMSVAKCFKKDITL 167
QY 160 AAPPFGNPJHNTMTIMRSVIDGILRYDTERDMAEDYQFWYDYSKI--GRLAYYPEALVKY 218
Db 168 KSPE-----KIKIDERLNYGDEVLFYIYFMFCCKIAVFKTCIYHY 208
QY 219 RLHAN---QVSSKHSVRQ-----HEIAOGIOKT-ARNDPLOSQ-----GFKTR 257
Db 209 EFNPNGRYENKKEILNQNHYDKKSNELIKLSKEFAHDEFQKLFLEVLRREAGVKNR 268

Search completed: December 2, 2002, 12:16:00
Job time : 8.97495 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 15.4445 Seconds
(without alignments)
2166.126 Million cell updates/sec

Title: US-10-007-267-3

Perfect score: 1823

Sequence: 1 MQLVSVLICAYNVEKYFAQ.....RLIKNRQARSDSACKDEI 348

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1712	93.9	346	2 H81970	lacto-N-neotetraos
2	1621.5	88.9	349	2 D81027	lacto-N-neotetraos
3	1617.5	88.7	333	1 S70813	glycosyl transfera
4	611.5	33.5	323	1 H64130	glycosyl transfera
5	327	17.9	367	2 G95948	probable glycosylt
6	293.5	16.1	333	2 B97168	glycosyltransferas
7	279.5	15.3	340	2 T44330	glycosyl transfera
8	276	15.1	321	2 AG2188	hypothetical prote
9	272	14.9	333	2 H97167	glycosyltransferas
10	271	14.9	336	2 A97168	glycosyltransferas
11	258	14.2	271	2 B84114	exopolysaccharide
12	252.5	13.9	343	2 A12091	glycosyltransferas
13	247	13.5	732	2 A84107	glycosyltransferas
14	245.5	13.5	318	2 AH2189	hypothetical prote
15	243	13.3	322	2 AE2160	glycosyltransferas
16	236.5	13.0	318	2 AG2189	hypothetical prote
17	235	12.9	333	2 AH2026	hypothetical prote
18	230.5	12.6	1013	2 AE1876	hypothetical prote
19	229.5	12.6	337	2 AG1920	hypothetical prote
20	229.5	12.6	696	2 D95206	hypothetical prote
21	228	12.5	278	2 D70036	exopolysaccharide
22	225	12.3	298	2 B75096	glycosyl transfera
23	224	12.3	290	2 H64431	glycosyl transfera
24	223.5	12.3	318	2 T50039	beta-1,4-galactosy
25	223	12.2	315	2 T44648	glycosyl transfera
26	219	12.0	334	1 G71153	glycosyl transfera
27	218	12.0	257	2 E84107	telchuronic acid b
28	217.5	11.9	294	2 E83022	probable glycosyl
29	216.5	11.9	281	2 G97777	glycosyl transfera

30	216.5	11.9	344	2 G70036	spore coat polysac
31	216	11.8	301	2 F95205	glycosyl transfera
32	214	11.7	280	2 E71703	glycosyl transfera
33	213	11.7	322	2 T44647	glycosyl transfera
34	210.5	11.5	344	2 A70037	capsular polysacch
35	208.5	11.4	389	2 E81318	probable galactosy
36	208.5	11.4	972	2 T09595	glucuronosyltransf
37	206.5	11.3	323	2 AD2189	hypothetical prote
38	205.5	11.3	311	2 T00087	rhmannosyltransfera
39	205	11.2	311	2 B83557	probable glycosyl
40	205	11.2	344	2 AC0974	probable glycosyl
41	204.5	11.2	250	2 A64099	glycosyl transfera
42	204	11.2	318	1 E71690	minor telchuric aci
43	203.5	11.2	327	2 AB1211	glycosyltransferas
44	203	11.1	316	2 AE2189	hypothetical prote
45	201.5	11.1	314	2 A71157	hypothetical prote

ALIGNMENTS

RESULT 1

H81970

lacto-N-neotetraose biosynthesis glycosyl transferase NMA0524 [Imported] - Neisseria

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: H81970

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491

A:Reference number: AB1775; MIMD:20222556; PMID:10761919

A:Accession: H81970

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-346 <PAR>

A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83816.1; PID:g737

A:Experimental source: serogroup A, strain 22491

C:Genetics:

A:Gene: lgtA; NMA0524

C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match	93.9%	Score 1712	DB 2	Length 346
Best local Similarity	94.3%	Pred. No. 1.6e-134		
Matches 328	Conservative 10	Mismatches 8	Indels 2	Gaps 1
QY 1	MQLVSVLICAYNVEKYFAQSLAAVYNQTRNDLIIYDDGSTDGLAIKDFQKRDSRI 60			
DB 1	MQLVSVLICAYNVEKYFAQSLAAVYNQTRNDLIIYDDGSTDGLAIKDFQKRDSRI 60			
QY 61	KILAQONSGILPSLNLINGIDELAKSGGGGGEYIARTDADDIASPGMTERKIVGEMEDRSI 120			
DB 61	KILAQONSGILPSLNLINGIDELAKS--GMGEYIARTDADDIASPMTERKIVGEMEDRSI 118			
QY 121	IANGAMLEVSEKEDGNRLARHHKGIWKKPTRHNDIAAFPPGNDIHNMTIMRSYI 180			
DB 119	IANGAMLEVSEKEDGNRLARHHKGIWKKPTRHNDIAADFFFGNPIHNMTIMRSYI 178			
QY 181	DGGLRDTFERMAEDYQFWYDVSKLGRFLAYYPALVKYRLHANOVSSKSVROHETAOGI 240			
DB 179	DGGLRDTFERMAEDYQFWYDVSKLGRFLAYYPALVKYRLHANOVSSKSVROHETAOGI 238			
QY 241	OKTARNDLOSMEKTRPDSLEYRQTKAAAYELPEKDIPDEPERARRFLYOQFKRTDPR 300			
DB 239	OKTARNDLOSMEKTRPDSLEYRQTKAAAYELPEKDIPDEPERARRFLYOQFKRTDPR 298			
QY 301	PSGAWDFADGRMRRLFTLRQYFGLILYRLIKNRQARSDSACKDEI 348			
DB 299	PSGAWDFADGRMRRLFTLRQYFGLILYRLIKNRQARSDSACKDEI 346			
RESULT 2				
DB1027				

lacto-N-neotetraose biosynthesis glycosyl transferase lgta NMB1929 [imported] - Neisseria

C:Species: Neisseria meningitidis

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: D81027

R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.

Hickey, E.K.; Halt, D.H.; Salberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;

ri, H.; Qiu, H.; Vamathavan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve

A>Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:2015755; PMID:10710307

A:Accession: D81027

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-349 <TEXT>

A:Cross-references: GB:AE002541; GB:AE002098; NID:g7227175; PID:AAF42258.1; PID:g722718

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB1929

C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match

Best Local Similarity 88.9%; Score 1621.5; DB 2; Length 349;

Matches 309; Conservative 14; Mismatches 10; Indels 1; Gaps 1;

QY 1 MQLPVSLICAYNVEKFAOSLAAVNQTWRNDILIVDGSITDGTALAKDFQKRSRI 60

DB 17 LQPLVSLICAYNVEKFAOSLAAVNQTWRNDILIVDGSITDGTALAKDFQKRSRI 76

QY 61 KILQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADIASPGWIEKIVGEMKDRSI 120

DB 77 RILQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADIASPGWIEKIVGEMKDRSI 135

QY 121 IAMGAMLEVLSEEDGNRLARRHKGKIMKPTREHEDIAAFPPGPNINNTMIMRSYV 180

DB 136 IAMGAMLEVLSEEDGNRLARRHKGKIMKPTREHEDIAAFPPGPNINNTMIMRSYV 195

QY 181 DGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSKSHVROHEIAOGI 240

DB 196 DGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSKSHVROHEIAOGI 255

QY 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEDEFERARFLYQCFKRTDP 300

DB 256 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEDEFERARFLYQCFKRTDP 315

QY 301 PSGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 334

DB 316 PAGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 349

RESULT 3

glycosyl transferase A (EC 2.4.-.-) - Neisseria meningitidis

C:Species: Neisseria meningitidis

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: S70813

R:Jennings, M.P.; Hood, D.W.; Peak, I.R.A.; Virji, M.; Moxon, E.R.

Mol. Microbiol. 18, 729-740, 1995

A>Title: Molecular analysis of a locus for the biosynthesis and phase-variable expressio

A:Reference number: S70812; MUID:96414473; PMID:8817494

A:Accession: S70813

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-333 <TEXT>

A:Cross-references: EMBL:U25839; NID:g973183; PID:AA04084.1; PID:g973185

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995

C:Genetics:

A:Gene: lgta

A:Start codon: TTG

C:Superfamily: Neisseria meningitidis glycosyl transferase A

C:Keywords: glycosyltransferase

Query Match

88.7%; Score 1617.5; DB 1; Length 333;

Best Local Similarity 92.5%; Pred. No. 1,1e-126;

Matches 309; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

QY 1 MQLPVSLICAYNVEKFAOSLAAVNQTWRNDILIVDGSITDGTALAKDFQKRSRI 60

DB 1 MQLPVSLICAYNVEKFAOSLAAVNQTWRNDILIVDGSITDGTALAKDFQKRSRI 60

QY 61 KILQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADIASPGWIEKIVGEMKDRSI 120

DB 61 KILQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADIASPGWIEKIVGEMKDRSI 119

QY 121 IAMGAMLEVLSEEDGNRLARRHKGKIMKPTREHEDIAAFPPGPNINNTMIMRSYV 180

DB 120 IAMGAMLEVLSEEDGNRLARRHKGKIMKPTREHEDIAAFPPGPNINNTMIMRSYV 179

QY 181 DGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSKSHVROHEIAOGI 240

DB 180 DGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSKSHVROHEIAOGI 239

QY 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEDEFERARFLYQCFKRTDP 300

DB 240 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEDEFERARFLYQCFKRTDP 299

QY 301 PSGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 334

DB 300 PAGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 333

RESULT 4

glycosyl transferase homolog H11578 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 26-Jul-2000

C:Accession: H64130

R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Weidman

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter

A>Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: H64130

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-323 <TEXT>

A:Cross-references: GB:U32832; GB:IA02023; NID:g1574421; PID:AA023227.1; PID:g1574422

C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match

Best Local Similarity 41.4%; Pred. No. 3.8e-43;

Matches 137; Conservative 62; Mismatches 113; Indels 19; Gaps 5;

QY 3 PLVSVLICAYNVEKFAOSLAAVNQTWRNDILIVDGSITDGTALAKDFQKRSRI 62

DB 5 PLVSVLICAYNVEKFAOSLAAVNQTWRNDILIVDGSITDGTALAKDFQKRSRI 64

QY 63 LQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADIASPGWIEKIVGEMKDRSI 122

DB 65 LQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADIASPGWIEKIVGEMKDRSI 118

QY 123 MGAMLEVLSEEDGNRLARRHKGKIMKPTREHEDIAAFPPGPNINNTMIMRSYV -D 181

DB 119 MGAMLEVLSEEDGNRLARRHKGKIMKPTREHEDIAAFPPGPNINNTMIMRSYV -D 177

QY 182 GGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSKSHVROHEIAOGI 241

DB 178 GGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSKSHVROHEIAOGI 237

QY 242 KTAARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEDEFERARFLYQCFKRTDP 301

DB 238 KTAARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEDEFERARFLYQCFKRTDP 288

QY 302 SGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 332

Db 289 ITS1LHFIK--YHLEFLDKQNLKIKKFR 317

RESULT 5

probable glycosyltransferase protein Smb21189 [imported] - *Sinorhizobium meliloti* (strain C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001 C:Accession: G95948

R.Finan, T.M.; Weiner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernan Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb *psymb* megaplasmid from the N2-fixing endo A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: G95948

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-367 <KUR>

A:Cross-references: GB:AL591985; PIDN:CAC49255.1; PID:g15140741; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid *psymb*

R.Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebaute, P.; Vandenbol, M.; Vorholter, F.J.; Weiner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: Smb21189

A:Genome: plasmid

Query Match 17.9%; Score 327; DB 2; Length 367; Best Local Similarity 29.5%; Pred. No. 2e-15; Matches 103; Conservative 68; Mismatches 128; Indels 50; Gaps 14;

3 PLVSLICAYNVEKYFAOSLAAYVNOTWRNLDLIVDGSSTDTGLAIKDFOKRDSRIKI 62

4 PVSIVLPYNAPEYIAAIESVLRODYERLEIADDSSTDSRILLEYKRSDSRVS 63

63 LAQONSGLIPSLNIGLDLAKSGGGGEYIARTDADIASPEWIEKIVGEMEKRSITIA 122

64 ISR-ENRGLATLNEGL-ALAK-----GELIARMADADIALPERSRLQVAFSAEPRLAL 116

123 MGAWLEVLSEKQGNRLARHHKGIKTKRPTREHEDIAAFPPGPNPHNNTMIRRSVI-D 181

117 SGGIDMLI---GNRIIR-GKPNPIY-RGSLRILISMFEET--FMHSTVYVNRNVIP 167

182 GGLRYDTERMDADYOFWYDVSRLGRAYPEALVRYRLHANOVSSKH-SVROHEIAOG 239

168 EMARVDPNVYHADFDLFRILADRPVHMIDELVAYRIHEDSVTSKHKRQMKRTILTY 227

240 IQKTARNDFLOSMGFKTRPDSLEYRQTKAAAYELP---EKDLEEDFERARR----- 288

228 AEMIALDALRLDLSAALGAIAVTSFVARLADVLALEREISAGGEYRRAVEDGALCF 287

289 --FLYGCFKRTDP-----PSGAWLDFADGRMRRLFTLRQYTGIL 327

288 EYFLYQLIAEEQPRRLTHEFLRTGKW-----GLIRR---RERYGL 326

RESULT 6

glycosyltransferase [imported] - *Clostridium acetobutylicum*
C:Species: *Clostridium acetobutylicum*
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B97168
R:Moiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*
A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK60133.1; PID:g15025169; GSPDB:GN00168
A:Experimental source: *Clostridium acetobutylicum* ATCC824
C:Genetics:
A:Gene: CAC2175

Query Match 16.1%; Score 293.5; DB 2; Length 333; Best Local Similarity 29.7%; Pred. No. 1e-16; Matches 80; Conservative 57; Mismatches 97; Indels 35; Gaps 10;

4 LVSVLICAYNVEKYFAOSLAAYVNOTWRNLDLIVDGSSTDTGLAIKDFOKRDSRIKI 63

1 MSIVMPYVNCERKYLESESTILKOTYRPFELIYVDSGNSIDINKYANDNRIVYV 60

64 AQAONSGLIPSLNIGLDLAKSGGGGEYIARTDADIASPEWIEKIVGEMEKRSITIA 123

61 SRDNMGWYVSLNEGIDR-AK-----GSYVARMDADIALPERFERQIEYLNKNKDVIDL 114

124 GAMLEVLSEKQGNRLARHHKGIKTKR--TRHEDIAAFPPGPNPHNNTMIRRSVID 181

115 ACKVEAFGVSRQKLERH-----WYNDLNNSEISLFLNCYIAHPVYMKKSVTLK 169

182 --GGLRYDTERMDADYOFWYDVSRLGRAYPEALVRYRLHANOVSSKH-SVROHEIAO 238

170 ALGGLVNLNR--TEDYINMLRLAIAGYITAMLEKTKIRLH-----NDSKIHROAE 220

239 GIQKTARNDFLOSMGFKTRPDSLEYRQTK 267

221 GF--SSIRDIQS-----RLEYVKEK 239

RESULT 7

glycosyl transferase homolog [imported] - *Vibrio cholerae*
C:Species: *Vibrio cholerae*
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44330

R.Yamasaki, S.; Shimizu, T.; Hoshino, K.; Ho, S.T.; Shimada, T.; Nait, G.B.; Takeda, Gene 237, 321-332, 1999

A:Title: The genes responsible for O-antigen synthesis of *Vibrio cholerae* 0139 are cl

A:Reference number: 222749; MUID:99453293; PMID:10521656

A:Accession: T44330

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-340 <YAM>

A:Cross-references: EMBL:AB012957; NID:g4115688; PIDN:BA43634.1; PID:g3721684

A:Experimental source: strain 022

C:Genetics:

A:Note: wblc

C:Superfamily: *Neisseria meningitidis* glycosyl transferase A

Query Match 15.3%; Score 279.5; DB 2; Length 340; Best Local Similarity 28.4%; Pred. No. 1.6e-15; Matches 82; Conservative 55; Mismatches 109; Indels 43; Gaps 9;

3 PLVSLICAYNVEKYFAOSLAAYVNOTWRNLDLIVDGSSTDTGLAIKDFOKRDSRIKI 62

6 PIVSVMSYNGEKYLAQAIESTILNOTFDFEITIVDDSDTSLSIIQAYMKDRIYL 65

63 LAQONSGLIPSLNIGLDLAKSGGGGEYIARTDADIASPEWIEKIVGEMEKRSITIA 122

66 ISRV-NKGLPYSLENAI-SVSKA-----NYIAMDDDISLPERETQALAVENNDICV 118

123 MGAWLEVLSEKQGNRLARHHKGIKTKRPTREHEDIAAFPPGPNPHNNTMIRRSVID 182

119 CGRLAVLFRFETSKNMCMH-----PEDHDSILITLLSVCIIRHVVMIRKSVLD- 168

183 GLRYDTERDM--AEDYQFVYDVSRLGRAYPEALVRYRLHANOVSSKH-SVROHEIAOG 240

169 QLDYVYENFRNSODYELMSRIAETFRFYTIQKPLLFYR-----DTPDGI 213

Db 231 TMK-AKIDYINDTNKKDKVDYL 251

RESULT 11

B84114
exopolysaccharide biosynthesis BH3714 [Imported] - *Bacillus halodurans* (strain C-125)
C:Species: *Bacillus halodurans*
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: B84114
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: B84114
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-271 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA807433.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3714

Query Match 14.2%; Score 258; DB 2; Length 271;
Best Local Similarity 31.5%; Pred. No. 7e-14;
Matches 64; Conservative 45; Mismatches 72; Indels 22; Gaps 7;

QY 1 MOPVSLICAVNVEKFFQSLAAVYVQWTRNDLIIYDGSIDGTLATAKDFQKRSRI 60
Db 1 MNPVSLVSVNDKNTLSLSTILNQTFFNEFLINDADSDSGELLEYSKDKRI 60

QY 61 KILAQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
Db 61 RLTHNNKNGLSYSLAEGV-SLAKA-----PWLRMDADVDVSKDLAQAOMVKAHSEL 114

QY 121 IAMGAWLEVLSEKDCGNRLARHHKHKIKKPTRHEDIAAF--PFGNFIHNTMIMRR 177
Db 115 DILGSY--VIDIDDKNEL-----EIRKVPPTHEIANLWTCPF---IHPTVLFKRD 162

QY 178 SVTDGLRYDTERDMAEDYQFWY 200
Db 163 SLTKAG-STDRNLRRQDYDLWF 184

RESULT 12

AI2091
glucosyltransferase [Imported] - *Nostoc* sp. (strain PCC 7120)
C:Species: *Nostoc* sp.
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AI2091
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakasaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Ana*
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2091
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-343 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA873987.1; PID:g17131380; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2288

Query Match 13.9%; Score 252.5; DB 2; Length 343;
Best Local Similarity 28.8%; Pred. No. 2.8e-13;
Matches 84; Conservative 63; Mismatches 100; Indels 45; Gaps 13;

QY 5 VSVLICAVNVEKFFQSLAAVYVQWTRNDLIIYDGSIDGTLATAKDFQKRSRIKILA 64
Db 3 LSVIIPVNSSESVAFELKRSVLAQYTRLEIITVDGSDSDKSIDICKOP--DERIRIIR 60

QY 65 QAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIANG 124
Db 61 Q-QNRLAGARNTGIRQ-----ARCEYLAFVSDDLMLPEKLAKLHEFRSEPEVSEF 113

QY 125 ANLEVLSEKDCGNRLARHHKHKK--TKKPTRHEDIAAFPP--GNPIIN-NTMIRRSV 179
Db 114 SRSLIDQ-----GKPLGIYQMP-KLTPIPEYLFGRNDPISNGSVVIRRAV 160

QY 180 ID-----GGRLRYDTERDMAEDYQFWYDV--KLGRLAYPEALVYKRLHANGV 225
Db 161 LDTIKQENMLYGEVEDFYFDSDFRQSEDIQEWRLALQTTWKTEGIPBALTYIRVMGGL 220

QY 226 SSKHSYRQHEIAQI--OKTARN-DFLOSGMFEKTRPDSLEYROTKAAYELP 274
Db 221 SA-NVLKQYESMRILLVKKQAVNPEFIERMGNNAKAYQLRYLARRTRRSP 271

RESULT 13

A84107
glucosyltransferase BH3657 [Imported] - *Bacillus halodurans* (strain C-125)
C:Species: *Bacillus halodurans*
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: A84107
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A84107
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-732 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA807376.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3657

Query Match 13.5%; Score 247; DB 2; Length 732;
Best Local Similarity 24.9%; Pred. No. 2.3e-12;
Matches 87; Conservative 59; Mismatches 114; Indels 90; Gaps 12;

QY 1 MOPVSLICAVNVEKFFQSLAAVYVQWTRNDLIIYDGSIDGTLATAKDFQKRSRI 60
Db 109 MEPKVTVIIIPAYNSEDEGIATFSIESIIQQTWTNIEIIVDSCSTDTALIKRYAASDDRY 168

QY 61 KILAQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
Db 169 KYLKTETNSGAVYARNIALQAQ--GEVYINDADWSHPAKETQVNHLEHPQV 222

QY 121 IAMGAWLEVLSEKDCGNRLARHHKHKIKKPTRHEDIAAFPPGPN-----IHNTMIM 175
Db 223 IANTSOQAMTEP-----LTFPRKGRGEXIYVNMSSLMF 257

QY 176 RRSVIDGLRYDTERDMAEDYQFWYVKL-GRLAYPEALVY--RLHANOVSSKHSV 231
Db 258 RKNPVAMENIGFMFVSVFAGDGEFRRLKRVFGEKAAVADLASGVSFQROTGSILTGNAF 317

QY 232 RQHEINQIOKTRARNDFLOSMGFEKTRPDSLE--YROTKAAYVLPKQD---LPEDFER 285
Db 318 GFHGYFMGVR-----KEYFDSYSHYRAESLRDFPOEPRFLFAVPEPMPQ 364

QY 286 ARREFLYQCFKRTDTPPSGAWLDFADGRMRRLFTLQYFGIL---YRLI 331
Db 365 RE-----KKT-----ADG-----RRHVDVITASFRLL 387

RESULT 14

AH2189
hypothetical protein alr3071 [Imported] - *Nostoc* sp. (strain PCC 7120)
C:Species: *Nostoc* sp.
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AH2189
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu

Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205–213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AH2189
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-318 <CUR>
 A:Cross-references: GB:BA000019; PIDN:BAE74770.1; PID:g17132165; GSPDB:GN00179
 A:Experimental source: strain FCC 7120
 C:Genetics:
 C:Gene: alr3071

Query Match	13.5%;	Score 245.5;	DB 2;	Length 318;
Best Local Similarity	28.9%;	Pred. No. 9.5e-13;		
Matches 70;	Conservative 45;	Mismatches 108;	Indels 19;	Gaps 5;

```

QY 3 PLSVLLCAYANEKYEFOSLAAVYVNOTRNLDLLIVDQSDGTSLAIKDKFQRDSRIKI 62
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 PKYSVLPAYINAMKYLPAVESVLOQSTETIELLIINGSSDN--IATMTAQLTDPROY 59

```

```

QY 63 LAQARNSGLIPSLNIGLDELAKSGGGGGEYIARTDADDIASPGWIEKIVEMEKDRSIIA 122
    ::|||::|: |||||::|::|::|:
Db 60 ISQ-QNQGSGARNCSIH-----ASEYIAFTDADDMLPTKEKQVCKLDNSPQAGL 112

```

```
QY      123  MGAMLEVLSEK-KDCGNRLARHHKHKGLTWKPTRHEDIAAFPEPCGNPIHNNTMIMRSVID 181
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      113  VYTWAMTDENKPTGCVIVASHHEGYWEQMVNDKIS-----NGSSAMVRICFD 163
```

QY 182 GGLRDTERRMAEDYFQWYDVKLGRALYYPEALVKYRLHANQVSSKHSVRQHEIQIOG 241
::: || | :: | | | : | :
Db 164 KVGLEDTLTSSEDRDMVIRLAHYHFAVVKEPITYLRHSQSMKKNRPKMLKNIRVFE 223

QY	242	KT	243
Db	224	KT	225

RESULT 15
AE2160
glycosyltransferase alr2836 [imported] - Nostoc sp. (strain PCC 7120)

A:Note: Nostroc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AE2160

Nakazaki, N., Shimo, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M., Tabata, S. DNA Res. 8, 205-213, 2001

A:Accession: AE2160
A:Status: preliminary
A:Molecule type: DNA

A:Cross-references: GB:BA000019; PIDN:BAW74535.1; PID:g17131930; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:

Query Match	13.3%;	Score 243;	DB 2;	Length 322;
Best Local Similarity	28.5%;	Pred. NO. 1.6e-12;		
Matches 96;	Conservative 61;	Mismatches 112;	Indels 68;	Gaps 19;

```
QY      5  VSVLICAYNVEKRYFAQSLAAVAVNQTRNLDILIVDDGSTDGTALAIAPQQR-DSRIKIL 63
      ::::  ||  :  ::::  ::  ||  :::::|||||||  :  ||:  :::  :
Db      3  ISVITSNRYARYLSRAINSVLAQTHSDLEIVIVDDGSTDNSRDVITQLOEQAPDKIKPI 62
```

```

QY      64 AQAQNSGLIPSLNIGHDELAKSGGGGEYIARTDADDIASPGWTEKIVGEMEKDRSIITAM 1233
        11 1 1 : 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db      63 FQA-NQGQGAFFNAGF-----AAATEVVALFDADDDVKKPHKIKRIV-LYFQTSDVGV 114

```

OY 124 GAMTEVLSEEDGN--RLARHHKGKIMKKPTRHEDIAAF-----PFGNPITHNNT 172
::: ||| : : | ||: | :

```

Db 115 MHHLDT-----DGNDKTIIDASTOG-----PKISEDLASVILDTGNAMCP-----PTSG 160
Qy 173 MIMRSVYIDGLRYDTE--RDMADYOFWYDVSKIGRLAYVPEALVYKRLHA--NOVSSK 228
   :  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:
Db 161 LAYRREYLEKEFPLDPYKWRITMA--DGLIITCTALGKIKTLOENLAVYRIHGANNHMSAA 219
Qy 229 HSVROHETAQ--GIQKTAR--NDFLQSMGKFTRPD---SLEYRQTKAAAYELPEKDLPE-- 280
   :  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:
Db 220 SATSEOEAKSQAGIEMTNOYINDLVLRIGARVDSLNRNLOYRRTK--YYORSQMDLREV 277
Qy 281 -----EDPERA---REFLVQCRK---RTD 298
   |||  |||:  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:  :  |||:
Db 278 WGISRLITGWPEFYGGQERAYVLAFLFKSGFELLRD 314

```

```
Search completed: December 2, 2002, 12:00:07
Job time : 17.4445 secs
```

```
Search completed: December 2, 2002, 12:00:07
Job time : 17.4445 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 8.46958 Seconds

(without alignments)
1704.189 Million cell updates/sec

Title: US-10-007-267-3
Perfect score: 1823
Sequence: 1 MQPLVSVLICAYNVEKYFAQ.....RLIKNRQARSASAKGEQEI 348

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	611.5	33.5	323	YF78_HAEIN	Q57287 haemophilus
2	243	13.3	322	YF86_ANASP	P22639 methanobacillus
3	229	12.6	290	YF57_METUA	Q58457 methanococcus
4	204.5	11.2	250	YF86_HAEIN	Q57022 haemophilus
5	189.5	10.4	344	YFBD_ECOCI	P11290 escherichia
6	187.5	10.3	266	YF86_HAEIN	Q46635 erwinia amy
7	164	9.0	301	YF86_HAEIN	Q46632 erwinia amy
8	163.5	9.0	446	YF86_HAEIN	P46917 bacillus su
9	161	8.8	900	YF86_HAEIN	P46918 bacillus su
10	146	8.0	267	YF86_HAEIN	Q48215 haemophilus
11	146	8.0	348	YF86_HAEIN	P33697 rhizobium m
12	144.5	7.9	268	YF86_HAEIN	P39614 bacillus su
13	143	7.8	256	YF86_HAEIN	P39621 bacillus su
14	135	7.4	441	YF86_HAEIN	P75905 escherichia
15	132.5	7.3	299	YF86_HAEIN	P75042 mycoplasma
16	132.5	7.3	909	YF86_HAEIN	P55465 rhizobium s
17	124.5	6.8	341	YF86_HAEIN	P75302 mycoplasma
18	122	6.7	342	YF86_HAEIN	P33700 rhizobium m
19	121.5	6.7	346	YF86_HAEIN	Q50587 mycobacteri
20	120.5	6.6	1275	YF86_HAEIN	Q50586 mycobacteri
21	120	6.6	299	YF86_HAEIN	P75086 mycobacteri
22	118.5	6.5	426	YF86_HAEIN	P42092 bacillus su
23	117.5	6.4	330	YF86_HAEIN	P33691 rhizobium m
24	116.5	6.4	424	YF86_HAEIN	P17862 rhizobium l
25	116.5	6.4	559	YF86_HAEIN	Q08912 mus musculu
26	114	6.3	297	YF86_HAEIN	P47306 mycoplasma
27	114	6.3	319	YF86_HAEIN	Q05090 mycobacteri
28	113.5	6.2	559	YF86_HAEIN	Q07537 bos taurus
29	113.5	6.2	559	YF86_HAEIN	Q10472 homo sapien
30	113.5	6.2	559	YF86_HAEIN	Q10473 rattus norv
31	112.5	6.2	256	YF86_HAEIN	Q50459 mycobacteri
32	111.5	6.1	323	YF86_HAEIN	P04319 bacillus su
33	111.5	6.1	426	YF86_HAEIN	P04341 rhizobium m

34	111.5	6.1	612	1	PAG3_CAEBL	P34678 caenorhabdi
35	110.5	6.1	298	1	Y025_MYCCE	P47271 mycoplasma
36	110.5	6.1	559	1	PAGE_PIG	Q29121 sus scrofa
37	109	6.0	289	1	RRBG_SHIFL	P37783 shigella fl
38	108	5.9	329	1	CSBB_BACSU	Q45539 bacillus fl
39	105.5	5.8	241	1	DPML_DROME	Q9Y1U7 drosophila
40	104.5	5.7	513	1	YTH1_RHOER	P46370 rhodococcus
41	103.5	5.7	236	1	DPML_SCHPO	Q14466 schizosacch
42	103.5	5.7	279	1	WCAA_ECOCI	P77414 escherichia
43	103	5.7	182	1	NODC_BRABL	P53417 bradyrhizob
44	98.5	5.4	260	1	DPML_HUMAN	Q60762 homo sapien
45	98	5.4	294	1	YF86_HAEIN	Q48214 haemophilus

ALIGNMENTS

```

RESULT 1
ID YF78_HAEIN STANDARD; PRT; 323 AA.
AC Q57287; 005077;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Putative glycosyl transferase H1578 (EC 2.-.-.-).
GN H1578.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=93550630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weisman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrichback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RL Rd.";
CC -f- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U32832; AAC23227.1; -.
DR TIGR: H1578; -.
DR InterPro: IPR001173; Glycosyltransferase_2.
DR Pfam: PF00535; Glycosyltransferase_2; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ
SEQUENCE 323 AA; 37680 MW; 7C8C2681039A5B4 CRC64;

Query Match 33.5%; Score 611.5; DB 1; Length 323;
Best Local Similarity 41.4%; Pred. No. 6.4e-43;
Matches 137; Conservative 62; Mismatches 113; Indels 19; Gaps 5;

OY 3 PLVSVLICAYNVEKYFAVAVVOTWNRDLIYDDSGTGLAIADFGKPSRIKI 62
DB 5 PLVSVVICAYNVEKYFAVAVVOTWNRDLIYDDSGTGLAIADFGKPSRIKI 64
OY 63 LAQANSGILPSINIGLDELAKSGGGGEYIARTADADIASPGWIEKIVGEMEKDRSIITA 122

```

```

Db      65 ISKRYMLGFINSLNIGL-----GCFSGKFAFAMADADDLAKPSWIKRIYVLEKKNHITA 118
      123 MGAMLEVLSEKGNRLARHHKHGKIMKRTREDIAAFPPENPIHNNMIRRSVIT-D 181
      119 MGSYLELIEVEKEGG-1IGSGYKTGIDMKPNLHNDICEAMLEFYNPIHNNMIRRANYRE 177
      182 GGLRYDTERDMADYQFWYDVSKLGRLAYVPEALVRYRLHANOVSCKSHVROHEIAQIG 241
      178 HKLFNKDYPYADKYFMSFVSRGLGLANPEALVRYRLHNGNOTSSVYNNQHEUTAKKIK 237
      242 KTRANDPLOSMTKPTPDSLEFYRTAKAAVEPEKDLPEEDFERARFLYQCKRRDTP 301
      238 RENTTYLKNIGIDIKVIN-----SVSLIEIYHVDKSNKVLKSLIEMYSIDKXT 288
      302 SGAMLEPDAADGRMRRLFTLRYFGILYRLIK 332
      289 ITSLLHFIR--YHLEFLDKONKITIKRIR 317

```

RESULT 2

YS86 ANASP

```

ID      YS86 ANASP      STANDARD:      PRT:      322 AA.
AC      P22639,
      01-AUG-1991 (rel. 19, Created)
      15-JUN-2002 (rel. 41, Last sequence update)
      15-JUN-2002 (rel. 41, Last annotation update)
      DE Putative glycosyl transferase alr2836 (EC 2.-.-.-).
      GN ALR2836
      OS Anabaena sp. (strain PCC 7120).
      OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
      OX NCBI_Taxid=103690;
      RN [1]
      RP SEQUENCE FROM N.A.
      RX MEDLINE=21595285; PubMed=11759840;
      RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
      RA Watanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,
      RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
      RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
      RA Yasuda M., Tabata S.;
      RA "Complete genomic sequence of the filamentous nitrogen-fixing
      RA cyanobacterium Anabaena sp. strain PCC 7120.";
      RL DNA Res. 8:205-213(2001).
      RN [2]
      RP SEQUENCE OF 1-131 FROM N.A.
      RX MEDLINE=90264305; PubMed=2111805;
      RA Holland D., Wolk C.P.;
      RA "Identification and characterization of beta, a gene that acts early
      RA in the process of morphological differentiation of heterocysts.";
      RL J. Bacteriol. 172:3131-3137(1990).
      CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: AF003591; BAB74535.1; -;
CC DR EMBL: AF031959; AAC32401.1; ALT_INIT.
CC DR PIR: B35391; B35391.
CC DR InterPro: IPR001173; Glycos_transf_2.
CC DR Pfam: PF00535; Glycos_transf_2; 1.
CC KW Hypothetical protein; Transferase; Glycosyltransferase;
CC Complete proteome.
CC SEQID 322 AA; 36388 MW; 75302PFB59327D968 CRC64;

```

```

Query Match      13.3%; Score 243; DB 1; Length 322;
Best Local Similarity 28.5%; Pred. No. 1e-12;
Matches 96; Conservative 61; Mismatches 112; Indels 68; Gaps 19;

```

```

QY      5 VSVLCAYNVEKYFAQSLAAYVNOTWRNLDILVDDSGTDLTAKDFQR-DSRIKIL 63
      3 ISVLIYNVYARLYSRALNSVLAQTHSDIEIYVDDSGTSDNSRDVITQEQAPDKIKPI 62
      64 AQONSGLPISLNTIGDELAKSGGGGEYIARTDADIASPGMIETIVEMEKDRSIIM 123
      63 FQA-NOGOGGAENAGF-----AAATGEVVAFLADADVAKPHKLQRIV-EVQTSDDVAV 114
      124 GAVLEVLSEKQDN--RLARHHKHGKIMKRTREDIAAF-----FPPGNPIHNNIT 172
      115 MHHLDIT-----DNDKTDIOASTQSG-----PKISEDLASVILQGTGNMCFP-----PTSG 160
      173 MIMRSVIDGLKRYDTE--RDMAEDYQFWYDVSKLGRLAYVPEALVRYRLHA--NOVSSK 228
      161 LAYRREVLEKVEPIDPVKKRIWA-DCGIITCYAFLGKIKITLQENLAYRYRHGANNHMSAA 219
      229 HSYRQHEINQ--GIQKTAR--NDPLOSMTKPTPDSLEFYRTAKAAVEPEKDLPE 280
      220 SATSEQAKSQAGIEENTNOYINDELVRIGYGARVDLSRNLQYRRK--YYORSQMDLREV 277
      281 -----EDFERA---RRFLYQCFK---RFD 298
      278 WGISRLILGMPFYSQGERAYLARFLFKSGKFLRSD 314

```

RESULT 3

YA57 METJA

```

ID      YA57 METJA      STANDARD:      PRT:      290 AA.
AC      058457,
      15-JUL-1998 (rel. 36, Created)
      30-MAY-2000 (rel. 39, Last sequence update)
      16-OCT-2001 (rel. 40, Last annotation update)
      DE Putative glycosyl transferase MJ1057 (EC 2.-.-.-).
      GN MJ1057.
      OS Methanococcus jannaschii.
      OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
      OC Methanocaldococcaceae; Methanocaldococcus.
      OX NCBI_Taxid=2190;
      RN [1]
      RP SEQUENCE FROM N.A.
      RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
      RX MEDLINE=96337999; PubMed=8688087;
      RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
      RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
      RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
      RA Overbeek R., Kirkness E.F., Weisslock K.G., Merrick J.M., Glodek A.,
      RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
      RA Uutterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
      RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
      RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
      RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
      RA jannaschii.";
      RL Science 273:1058-1073(1996).
      CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: U67549; AAB99061.1; -.
CC DR HSSP: P39621; IQGQ.
CC DR TIGR: MJ1057; -.
CC DR InterPro: IPR001173; Glycos_transf_2.
CC DR Pfam: PF00535; Glycos_transf_2; 1.
CC KW Hypothetical protein; Transferase; Complete proteome.
CC SEQID 290 AA; 35099 MW; 3f6A1B221C420D74 CRC64;

```

```

Query Match      12.6%; Score 229; DB 1; Length 290;
Best Local Similarity 29.9%; Pred. No. 1.3e-11;

```



```
Dh 56 ERYSGL--KINLVISETKINGAGARRNCIDLATGDYVCFDADD---EWHKDLQCMLS 109
Oy 107 -IERKIVGEMEK-----DRSIIAMGAMLEVLSEEDKGRRLARHHKHGKIMKKPIFHEDIA 159
Dh 110 LIEREGOGDRRIIITYSVNIIQDSFLKMPKLVG-----EHESIA 152
Oy 160 AAFPFG--NP1HNMTMIMRSVYDGG--LRYDTERDMEDYQFWYDVSFKIGRLAYPEALV 216
Dh 153 EYL-EGCYGFIQTSTYVLRK--DAAEIRFDERYIRHDDYLCIRADLG-----F 200
Oy 217 KYRLHANOVSSKHSYRQHEIAGIOKTKARNDPLOSNGCKTFRPDSLEYRQOTAAAYELPEK 276
Dh 201 KFMIAIADPLANYHMV-----TRFGS--QHKGSYKVSFLFWL 234
Oy 277 DLPEDEFARARFLYOCFKRTDTPPSGAMLPADGR-MRLSLFTLROYE 324
Dh 235 DAMPHLTRDYVITYYKAKLP-----LKKMDKSLQSLSPARYF 275

RESULT 8
GGAB_BACSU STANDARD; PRT; 446 AA.
AC P46917;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
Df Minor telchoic acids biosynthesis protein ggaA.
GN GGA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Freymond P., Karamata D.;
RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boursier L., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Delozot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Ertlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.W., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Portwoll S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivoita C., Roche E., Roche B., Rose M., Sadle Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Yamakoshi A., Tanaka T., Terpestra P., Toponoti A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler H., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF GALACTOSAMINE-CONTAINING
CC MINOR TELCHOIC ACIDS.
```

```
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U13979; AAA73512.1; -.
CC EMBL: Z99122; CAB15586.1; -.
CC Subtilisin; BG1367; ggaA.
CC InterPro: IPR001173; Glycos. transf. 2.
CC Pfam: PF00535; Glycos. transf. 2; 1.
CC Transferrase; Glycosyltransferase; Complete proteome.
CC SEQUENCE 446 AA; 53148 MW; BB3698D96856C42 CRC64;
```

```
Query Match 9.0%; Score 163.5; DB 1; Length 446;
Best Local Similarity 19.8%; Pred. No. 5.2e-06;
Matches 74; Conservative 72; Mismatches 139; Indels 89; Gaps 15;
```

```
Oy 4 LVSYLICAYNWEKYPFAGSLAAVYNOT---WRNLDILYDGSSTDTTIAIADFQK-RDSR 59
Dh 1 MFSIIIPYNSENDYLRYSIESVLTNOSIGFKEINELIILDDGSVDSPQICESFKMLYPNN 60
Oy 60 IKILAQANSGILPSNLGLDELAKSGGGGEYIARTADDIASGWTIEKIVGEMEKRS 119
Dh 61 IKIM-KIENSGPSAARNGLSNVSR---SKFIFLSDSDAFSONALQSYV-DEFCSE 114
Oy 120 IIAMGAMLEVLSEEDKGRRLARHHKHGKIMKKPIFHEDI---AAFPFGNP1HNMTM 174
Dh 115 HVNIAVLPLVYTYGKEGG---HKLNRFKFGTRVIMILDYRIHYITGTF----- 164
Oy 175 MRSYIDGLRYDTERDMEDYQFWYDVSFKIGRLAYPEALVKYRLHANOVSSKHSYRQ 234
Dh 165 -RRHTLTSTVLFD-----ESIKWEIDA-----IFFNQLLKEKRYGAVAGKYYFKR 211
Oy 235 ELAAGIOKTA-----RNFELOSMGKFT-----RFDLS-----EYR 264
Dh 212 KEQDSLVDRSFNNKRYTYLLNECYMTLMDSFNNKYDVLVPLQFLIYHKLFLYPNR 271
Oy 265 QTKAAAYELPKEDLPEDFERARFL-YQCFKRTDTPPSGAMLPADGRMRRLFTLROY 323
Dh 272 DVYSVLDQOGQRAVVDFFIVKIKIDQFIKEQDM-----MYT 311
Oy 324 FGLYRLIKNRQA 337
Dh 312 KEFMHLLKENTEA 325

RESULT 9
GGAB_BACSU STANDARD; PRT; 900 AA.
AC P46918;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
Df Minor telchoic acids biosynthesis protein ggaB.
GN GGA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Freymond P., Karamata D.;
RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
```

```

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codant J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Jones L.,
RA Hilbert H., Holstapp S., Hosono S., Hull J.M., Itaya M., Jones A.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Koyashiki Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Melado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Patro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presean E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF GALACTOSAMINE-CONTAINING
CC MINOR TETRAHIC ACIDS.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL: U13979; AAA73513.1; -.
DR EMBL: Z99122; CAB15585.1; -.
DR Subtilisin; BG1192; 99ab.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Transferase; Glycosyltransferase; Complete proteome.
SQ
SEQUENCE 900 AA; 107154 MW; FA66495488C2C62F CRC64;

Query Match      8.8%; Score 161; DB 1; Length 900;
Best Local Similarity 22.8%; Pred. No. 2.2e-05;
Matches 75; Conservative 59; Mismatches 137; Indels 58; Gaps 13;

```

```

OY 6 SVLICAYNEKFEAOSLAAYVNOT--WRNLDILYDDGSTDGTALA-IAKDFOKRDSRIKL 63
DB 11 SVIMPIYNEVLITFAISIIINOTIGFENIOILIVNDSDPDSSEITCKYACKYNNNIAY 70
OY 64 AOAONSGILPSINIGLDELAKSGGGGEYIARTDADDIASPGWIKIYGEKMSIIAM 123
DB 71 AKKONGVSSANNYGKTY-----AEGRTIOFLDPDDLVSECTFEVLFNFEHKNKEIDI 124
OY 124 GAMLEVLSEKDNRLARHKGKTIWKPTRHEDIAAFPPGNPIH-----NNTMIIMRS 178
DB 125 VAIRPFFPAGRG-----EHNLNKPF--SSTRILDVEK-----EPHIIILHOCSTFIKKA 173
OY 179 VIDGILRTDEMDADYQFW-YDVSKLGRILAYPEALVKKYRLH-----ANQVS 226
DB 174 L--KNIRFENCKIGEDAKLVNLIISOKKKYGLVKEAKYHYVRREDGSSAMQTAANKMW 231
OY 227 SKHS-----VRQHE--IAOGIQKTANDPLQSKGFR---TRDSLEYKOTKA 269
DB 232 FNHSILTFESKNLIDIKNEOKIPFLQYVMVMDKWKLLIKDISETPLDENESYSEFTL 291

RESULT 10
YG95_HAEIN STANDARD; PRT; 267 AA.
ID YG95_HAEIN
AC 048215; 005081;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase H11695 (EC 2.2.1.1).
GN H11695.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=A2;
RA McLaughlin R., Abu Kwaik Y., Young R., Spinola S., Apicella M.;
RT "Characterization and sequence of the usg locus from Haemophilus
RT influenzae.";
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Furmann J.L., Georgiagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL: M94855; AAA24983.1; -.
DR EMBL: U32842; AAC23341.1; -.
DR TIGR; H11695; -.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
FT CONFLICT 26 V -> G (IN REF. 1).
FT CONFLICT 46 D -> E (IN REF. 1).
FT CONFLICT 49 F -> S (IN REF. 1).
SQ
SEQUENCE 267 AA; 30770 MW; A2F1A0532737D8C3 CRC64;

Query Match      8.0%; Score 146; DB 1; Length 267;
Best Local Similarity 24.3%; Pred. No. 7.2e-05;
Matches 74; Conservative 47; Mismatches 129; Indels 54; Gaps 11;

```

```

Db      63  YALPQNRGKALNGLHC-----DYDWVFRMDTDDICVDREKQVAFIEQPESTII 116
Oy      123  MGAMLEVLSEEDNGRLRHHKHKIMKPTRHEDIAFFPPGNGIHNNTMTMRSDVDG 182
Db      117  FGG--QIAEFEGNNVDIAVYR-----NVPTSAQELIKFTQRCRPFNHTVAYQKSAVIN 168
Oy      183  GLRYDTERDMAEDYQFWYDVSKLG-RLAYYPEALVKYRLHANQVSKHSVRQH----- 234
Db      169  CGGE---DLQEDDYLTWKIVAGLYMANLPDILYARVNGMVRGRCVNGAKAEMRLF 225
Oy      235  --EIAOGIOKTARNDPLOSMEGFKTFPDSLEYRQTKAAAYELPEKDLPEDEFERARFLYQ 292
Db      226  KUKYRLGIQGLSLGSLFALRFGSRL-----LPTSLDKK-----LYQ 262
Oy      293  CFKR 296
Db      263  TFLR 266

RESULT 11
EXOO_RHIME
ID      EXOO_RHIME      STANDARD:      PRT:      348 AA.
AC      P33697.
DC      01-FEB-1994 (Rel. 28, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Succinoglycan biosynthesis protein exoo (EC 2.-.-.-).
GN      EXOO OR Rbl084 OR SWS20959.
OS      Rhizobium meliloti (Sinorhizobium meliloti).
OC      Bacteria; PSymb (megaplasmid 2).
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX      NCBI_TaxID=382;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=1021;
RX      MEDLINE=94042869; PubMed=8226645;
RA      Glucksmann M.A., Reuber T.L., Walker G.C.;
RT      "Family of glycosyl transferases needed for the synthesis of
RL      succinoglycan by Rhizobium meliloti."
RN      J. Bacteriol. 175:7033-7044(1993).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=RCR2011 / SU47;
RX      MEDLINE=94067019; PubMed=8246891;
RA      Becker A., Kleickmann A., Keller M., Arnold W., Puehler A.;
RT      "Identification and analysis of the Rhizobium meliloti exoAMONP genes
RT      involved in exopolysaccharide biosynthesis and mapping of promoters
RT      located on the exoHKLAMONP fragment."
RL      Mol. Gen. Genet. 241:367-379(1993).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=1021;
RX      MEDLINE=21396508; PubMed=11481431;
RA      Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA      Vorholster F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA      Golding B., Puehler A.;
RT      "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT      fixing endosymbiont Sinorhizobium meliloti."
RL      Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC      -I- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF
CC      SUCCINOGLYCAN (EPS I). NEEDED FOR THE ADDITION OF THE FIFTH SUGAR
CC      (GLUCOSE), CATALYZES THE FORMATION OF A BETA-1,6 LINKAGE BETWEEN
CC      THE FOURTH AND FIFTH SUGAR.
CC      -I- PATHWAY: Exopolysaccharide biosynthesis.
CC      -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC      -I- CAUTION: Ref.1 sequence differs from that shown due to a
CC      frameshift in position 208.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed, usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L20758; AAA16044.1; ALT_FRAME.
DR      EMBL; 226336; CAA80347.1; -.
DR      EMBL; AL603645; CACA9484.1; -.
DR      PIR; C49348; C49348.
DR      PIR; S37029; S37029.
DR      PIR; S3958; S3958.
DR      InterPro; IPR001173; Glycos_transf_2.
DR      Pfam; PF00535; Glycosyltransferase; Exopolysaccharide synthesis;
KW      Transferase; Glycosyltransferase; Exopolysaccharide synthesis;
KW      Plasmid; Complete proteome.
FT      DOMAIN 38 45
FT      SEQUENCE 348 AA; 38131 MW; 770B43782F785579 CRC64;
SQ
Query Match      8.0%; Score 146; DB 1; Length 348;
Best local similarity 23.7%; Pred. No. 0.0001;
Matches 81; Conservative 51; Mismatches 144; Indels 66; Gaps 14;
Oy      3  PLVSVLICAVNYEKEYFAOSLAAVNQWTRNIDILIVDDGSTDGLAIKDFQKRDRIKI 62
Db      10  PVTEFVAAYNASDITVRIAESALAQEGVTEVVVDDCSAATFALVAAP--DPRVRL 67
Oy      63  LQAQNSGLIPSLNIGLDELAKSGGGGGEYIARTADDIASGGMTEKYEKEKRSITIA 122
Db      68  IALDRNRGPGGARNNGI-----GAARGFWIAVLSDSDPRDRLRMIERADAGAOTA 121
Oy      123  MGAMLEVLSEEDNGRLRHHKHKIMKPTRHEDIAF-----PPFGNPITHN 170
Db      122  VDN-LDVS--LDGRSL-RMFESEALARP--QILPAFIESNVLPFRSHNNGYKPIFE 175
Oy      171  NTMIRRSYIDGLRYDTERDMAEDYQFWYDVSKL--GRLAYYPEALVKYRLHANQVSS 227
Db      176  -----RRFLENOQLRFDEALRIGEDYILL--ASALACGGRCAPERSA--GYIYHIREGSI 226
Oy      228  KISVYQHLEAOGIOKTARNDPLOSMEGFKTFPDSLEYRQTKAAAYELPEKDLPEDEFERAR 287
Db      227  SKVLRDHDIDAMI--AADEAFLLRRYALDGLAQKMRHRRRG-----FREAR 270
Oy      288  RFLYOCFKRTDTPPSGAWLDFAAD-----GRMRRL 317
Db      271  SFLVLEQLKRSKSLAGLTKTALADFPALRHLSMPTAARLRL 312

RESULT 12
YWDF_BACSU
ID      YWDF_BACSU      STANDARD:      PRT:      268 AA.
AC      P39614.
DC      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Putative glycosyl transferase ywdf (EC 2.-.-.-).
GN      YWDF OR IPA-56D
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=1423;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168;
RX      MEDLINE=95020537; PubMed=7934828;
RA      Glaser P., Kunst F., Arnaut M., Coudart M.P., Gonzales W.,
RA      Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA      Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA      Rapoport G., Danchin A.;
RT      "Bacillus subtilis genome project: cloning and sequencing of the 97
RT      kb region from 325 degrees to 333 degrees."
RL      Mol. Microbiol. 10:371-384(1993).
RN      [2]
RP      SEQUENCE FROM N.A.

```

```

RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bortier M.G., Bessieres P., Bolotin A., Borchert S.,
RA Brouillet S., Bruschier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Choi S.K., Codani J.J., Comercon I.F., Cummings N.J., Daniel R.A.,
RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RT Nature 390:249-256(1997).
CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL: X73124; CAA51612.1; -
CC DR EMBL: 269123; CAB15824.1; -
CC DR PIR: S39711; S39711.
CC DR Subtilist: Bg10602; ywdf.
CC DR InterPro: IPR001173; Glycos_transf_2.
CC DR Pfam: PF00535; Glycos_transf_2; 1.
CC KW Hypothetical protein; transferase; glycosyltransferase;
CC Complete proteome.
CC SEQUENCE 268 AA; 30616 MW; DB8428F7016EC9B3 CRC64;

```

```

OY 219 -- RLHANOVSRSKRV 231
DB 214 KRRQYLDRLSEKHL 228
RESULT 13
SESA_BACSU STANDARD: PRT; 256 AA.
ID SPSA_BACSU
AC 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Spore coat polysaccharide biosynthesis protein spsa.
GN SPSA OR IPA-63D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaut M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Verres A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees."
RL Mol. Microbiol. 10:371-384(1993).
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bortier M.G., Bessieres P., Bolotin A., Borchert S.,
RA Brouillet S., Bruschier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Choi S.K., Codani J.J., Comercon I.F., Cummings N.J., Daniel R.A.,
RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RT Nature 390:249-256(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX MEDLINE=99280694; PubMed=10350455;
RA Charnock S.J., Davies G.J.;
RT "Structure of the nucleotide-diphospho-sugar transferase, SpsA from
RT Bacillus subtilis, in native and nucleotide-complexed forms."
CC Biochemistry 38:6380-6385(1999).
CC -I- FUNCTION: GLYCOSYLTRANSFERASE IMPLICATED IN THE SYNTHESIS OF THE
CC SPORE COAT.
CC -I- PATHWAY: Spore coat polysaccharide biosynthesis.

```

```

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X73124; CAA51619.1; -
DR EMBL: 299123; CAB15817.1; -
DR PDB: 1OG8; 21-APR-00.
DR PDB: 1OG8; 04-MAY-00.
DR PDB: 1OGS; 04-MAY-00.
DR Subtilisin: Bg10609; spsa.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_1.
DR Transferrin: Glycosyltransferase; 3D-structure; Complete proteome.
DR Disulfid 155 243
SQ SEQUENCE 256 AA; 30184 MW; C20EA9627F5D536B CRC64;

Query Match 7.8%; Score 143; DB 1; Length 256;
Best Local Similarity 22.9%; Pred. No. 0.00012;
Matches 62; Conservative 66; Mismatches 87; Indels 56; Gaps 13;

OY 3 PLVSVLICAYNEKYPAOISLAAVNONTWNRDLIYDGSSTGTLIAADFOKRSRIKI 62
DB 2 PKVSVMTSYNKSVDYAKAISLSLSTQSFDFELTMDNSNETLAVIRPF-LNDNRKRF 60
OY 63 LAOQNSGL-----IPSLNIGDELAKSGGGGGEYIARTDADDIASPGWIEKIVE 113
DB 61 Y-QSDISGVKERTKRYVALINQAI-EMAE-----GEYITATDNTNIMPDLLKMYRE 113
OY 114 ME-KRSTIIMGAMLEVISEKDGRLARHHKHKIKKPRRHEDIAAFPPGPN--IH 169
DB 114 LTHPKAVIYASAKTYHLNEND-----YKREYAP--AAQVTWNAACAD 158
OY 170 NMTMIRRSVIDGLRYPDERDMAE-----DYQFWYDVSGLRLAYPEALVYRLH 221
DB 159 HCSVMIRYSVLE-KVKEKGSYWDSPAFRYIGDARFFRVNH-----FYPPFLDEILD 212
OY 222 ANOVSSKHSVRHQLAOGIQTARNDFLOSM 252
DB 213 LNVITDQ-----SIHQFLFELEKNEFVRL 237

RESULT 14
YCDQ_ECOLI
ID YCDQ_ECOLI STANDARD: PRT; 441 AA.
AC P75905;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ycdQ.
GN YCDQ OR B1022.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Grou B., Shao Y.,
RA Mau G., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RT "The complete genome sequence of Escherichia coli K-12.";
RN [2]
RP Science 277:1453-1474(1997).
RC STRAIN-K12;

```

```

RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Samped G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RN DNA Res. 3:137-155(1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: TO Y.PESTIS HNS LOCUS PROTEIN HNSR AND TO
CC S.EPIDERMIDIS ICA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000204; AAC74107.1; -
DR EMBL: D90739; BAA35803.1; -
DR EcoGene: EG13863; ycdQ.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_1.
DR Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 5 25
FT TRANSMEM 33 53
FT TRANSMEM 331 351
FT TRANSMEM 363 383
FT TRANSMEM 395 415
SQ SEQUENCE 441 AA; 50765 MW; 04F5A53D72FEBABB CRC64;

Query Match 7.4%; Score 135; DB 1; Length 441;
Best Local Similarity 23.9%; Pred. No. 0.0011;
Matches 57; Conservative 35; Mismatches 86; Indels 60; Gaps 9;

OY 3 PLVSVLICAYNEKYPAOISLAAVNONTWNRDLIYDGSSTGTLIAADFOKRSRIKI 62
DB 75 PSTSIITPCFNEKNEKVEITHALAQRYENIEIVANDSTGTRAILDPMQAQIPHLRF 134
OY 63 LAOQNSGLIPSLNIGDELAKSGGGGGEYIARTDAD-----IASPGWIEKIVE 113
DB 135 IHLAQNGKAIKATGA-AAAKS-----EYLCIDGDALLDRAAYIYEPHMYNRVGA 188
OY 114 MEKRSI-----IANGAMLEVISEKDGRLARHHKHKIKKPRRHEDIAAFPPF 164
DB 189 VTGNPRIRTRSTLVKRIQGEYSIILIKRTQRI-----YGVNF--TVSGVIAAF-- 237
OY 165 GNPIDHNMTIMRSVY-----DGLRYPDERDMAEDYQFWYDVSGLRLAYPEAL 215
DB 238 -----RRSALAEVGSWSDMTEDIDISWKLQINQW-----TIFYEPRAL 277

RESULT 15
Y060_MYCPN
ID Y060_MYCPN STANDARD: PRT; 299 AA.
AC P75042;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase MG060 homolog (EC 2.-.-.-)
GN (D09_ort299).
GN MPN075 OR MP080.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RC MEDLINE=97105885; PubMed=8948633;

```

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 : Search time 29.8926 Seconds
(without alignments)
2398.732 Million cell updates/sec

Title: US-10-007-267-3
Perfect score: 1823
Sequence: 1 MQLPLSVLICAVNVEKYFAQ.....RLIKNRQARSASGKEGEI 348

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_minc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1823	100.0	348	2	Q50946
2	1788	98.1	346	2	Q9RGN2
3	1785	97.9	362	2	Q50951
4	1773	97.3	346	2	Q93PS0
5	1712	93.9	346	16	Q9JW66
6	1657.5	90.9	322	2	Q918F0
7	1651	90.6	321	2	Q918E9
8	1621.5	88.9	349	16	Q9JX06
9	1617.5	88.7	333	2	Q51115
10	1609	88.3	348	2	Q9EVD5
11	1118	61.3	221	2	Q93EK6
12	1049	57.5	337	2	Q50949
13	666.5	36.6	337	16	Q9CIE9
14	412.5	22.6	337	2	P74947
15	410	22.5	337	2	Q34234
16	373	20.5	172	2	Q9RGM9

17	346	19.0	149	2	P96946	P96946 neisseria m
18	327	17.9	367	16	Q92V61	Q92V61 rhizobium m
19	318	17.4	80	2	Q9K2R3	Q9K2R3 neisseria g
20	312	17.1	368	16	Q98JH2	Q98JH2 rhizobium 1
21	303	16.6	117	2	P96944	P96944 neisseria m
22	300	16.5	133	2	Q93EK8	Q93EK8 neisseria m
23	293.5	16.1	333	16	Q97H38	Q97H38 clostridium
24	279.5	15.3	340	2	Q87159	Q87159 vibrio chol
25	276	15.1	321	16	Q8XSW2	Q8XSW2 anabaena sp
26	274	15.0	288	2	Q56869	Q56869 yerstinla en
27	272	14.9	333	16	Q97H40	Q97H40 clostridium
28	271	14.9	336	16	Q97H39	Q97H39 clostridium
29	267.5	14.7	314	2	Q91A88	Q91A88 aeromonas h
30	258	14.2	271	16	Q9K615	Q9K615 bacillus ha
31	252.5	13.9	343	16	Q8YUP7	Q8YUP7 anabaena sp
32	247	13.5	732	16	Q9K650	Q9K650 bacillus ha
33	245.5	13.5	318	16	Q8YSL3	Q8YSL3 anabaena sp
34	236.5	13.0	318	16	Q8YSL4	Q8YSL4 anabaena sp
35	235	12.9	333	16	Q8YWS0	Q8YWS0 anabaena sp
36	234.5	12.9	354	2	Q9AH91	Q9AH91 streptococc
37	230.5	12.6	1013	16	Q8YVZ1	Q8YVZ1 anabaena sp
38	229.5	12.6	337	16	Q8YVD6	Q8YVD6 anabaena sp
39	229.5	12.6	696	16	Q97P72	Q97P72 streptococc
40	228.5	12.5	306	2	Q9RP62	Q9RP62 escherichia
41	228	12.5	257	2	Q93CQ9	Q93CQ9 shigella bo
42	228	12.5	278	16	P71054	P71054 bacillus su
43	227	12.5	279	2	Q9RG44	Q9RG44 streptococc
44	225	12.3	298	17	Q9UZ16	Q9UZ16 pyrococcus
45	224	12.3	301	2	Q9S410	Q9S410 leptospira

ALIGNMENTS

RESULT 1	
Q50946	PRELIMINARY; PRT; 348 AA.
AC	Q50946; 01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	Glycosyl transferase.
GN	LGTA.
OS	Neisseria gonorrhoeae.
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX	NCBI_TaxID=485;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=F62;
RX	MEDLINE=95053752; PubMed=7964493;
RA	Gotschlich E.C.;
RT	"Genetic locus for the biosynthesis of the variable portion of
RT	Neisseria gonorrhoeae lipooligosaccharide.";
RL	J. Exp. Med. 180:2181-2190(1994).
DR	EMBL; U14554; AAA68009.1;
DR	InterPro: IPR001173; Glycos_transf_2.
DR	Pfam: PF00535; Glycos_transf_2; 1.
KW	Transferase.
SO	SEQUENCE 348 AA; 39975 MW; ECBBF2FACD84C72A CRC64;
Query Match	100.0%; Score 1823; DB 2; Length 348;
Best local Similarity	100.0%; Pred. NO. 1.5e+148;
Matches 348;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MQLPLSVLICAVNVEKYFAQSLAAVYNQTRNLDTIYVDGSTDGLAIKDFQKRDSRI 60
DB	1 MQLPLSVLICAVNVEKYFAQSLAAVYNQTRNLDTIYVDGSTDGLAIKDFQKRDSRI 60
QY	61 KTLAAQNSGLIPSNIGLDELAKSGGGGEYIARPDADDIASPGWIEYIVGEMEDRSI 120
DB	61 KTLAAQNSGLIPSNIGLDELAKSGGGGEYIARPDADDIASPGWIEYIVGEMEDRSI 120
QY	121 IAMGAWLEVLSEKDKGNRLARRHHKGGIKWKPTRHEDIAAFPFGNPJHNNTMIMRSYI 180

```

DB 121 IAMGAWLEVLSEKQGNRLARHHKHKIMKKPTRHEDIAPFPFGNPINNNTMIMRSYI 180
      |||
QY 181 DGGLRVDTFRDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHSVRQHEIAGOI 240
      |||
DB 181 DGGLRVDTFRDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHSVRQHEIAGOI 240
      |||
QY 241 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
      |||
DB 241 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
      |||
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
      |||
DB 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
      |||

RESULT 2
QYRGN2 ID QYRGN2 PRELIMINARY; PRT; 346 AA.
AC QYRGN2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Glycosyltransferase.
GN LGTA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1291;
RX MEDLINE=20305049; PubMed=10844691;
RA Harvey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
  Phillips N.J., Apicella M.A., Balke M.S.;
  "Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein
  receptor on human sperm.";
  Mol. Microbiol. 36:1059-1070(2000).
RL EMBL; AF121135; AAF14359.1; -.
DR InterPro: IPR001173; Glycos. transf_2.
DR Pfam: PF00535; Glycos. transf_2; 1.
KW Transferase.
SQ SEQUENCE 346 AA; 39730 MW; A7E9434C4635815D CRC64;

```

```

Query Match 98.1%; Score 1788; DB 2; Length 346;
Best Local Similarity 98.6%; Pred. No. 1.5e-145;
Matches 343; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

```

```

QY 1 MOPVSVLICAVNVEKYFAQSLAAVYNQWNRNLDILIVDGSIDGTGLAIKDFQKRDSTI 60
      |||
DB 1 MOPVSVLICAVNVEKYFAQSLAAVYNQWNRNLDILIVDGSIDGTGLAIKDFQKRDSTI 60
      |||
QY 61 KIQAQONSGILPSLNINGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
      |||
DB 61 KIQAQONSGILPSLNINGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 118
      |||
QY 121 IAMGAWLEVLSEKQGNRLARHHKHKIMKKPTRHEDIAPFPFGNPINNNTMIMRSYI 180
      |||
DB 119 IAMGAWLEVLSEKQGNRLARHHKHKIMKKPTRHEDIAPFPFGNPINNNTMIMRSYI 178
      |||
QY 181 DGGLRVDTFRDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHSVRQHEIAGOI 240
      |||
DB 179 DGGLRVDTFRDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHSVRQHEIAGOI 238
      |||
QY 241 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
      |||
DB 239 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 298
      |||
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
      |||
DB 299 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 346
      |||

```

RESULT 3

```

QY0951 ID QY0951 PRELIMINARY; PRT; 362 AA.
AC QY0951;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE Glycosyltransferase.
GN LSI2.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1291;
RX MEDLINE=96099316; PubMed=8522539;
RA Danaher R.J., Levin J.C., Arking D., Burch C.L., Sandlin R.,
  Stein D.C.;
  "Genetic basis of Neisseria gonorrhoeae lipooligosaccharide antigenic
  variation.";
  J. Bacteriol. 177:7275-7279(1995).
RL EMBL; U15992; AAA92074.1; -.
DR InterPro: IPR001173; Glycos. transf_2.
DR Pfam: PF00535; Glycos. transf_2; 1.
KW Transferase.
SQ SEQUENCE 362 AA; 41623 MW; ED67B468C9E3C3A CRC64;

```

```

Query Match 97.9%; Score 1785; DB 2; Length 362;
Best Local Similarity 98.3%; Pred. No. 3e-145;
Matches 342; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

```

```

QY 1 MOPVSVLICAVNVEKYFAQSLAAVYNQWNRNLDILIVDGSIDGTGLAIKDFQKRDSTI 60
      |||
DB 17 MOPVSVLICAVNVEKYFAQSLAAVYNQWNRNLDILIVDGSIDGTGLAIKDFQKRDSTI 76
      |||
QY 61 KIQAQONSGILPSLNINGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
      |||
DB 77 KIQAQONSGILPSLNINGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 134
      |||
QY 121 IAMGAWLEVLSEKQGNRLARHHKHKIMKKPTRHEDIAPFPFGNPINNNTMIMRSYI 180
      |||
DB 135 IAMGAWLEVLSEKQGNRLARHHKHKIMKKPTRHEDIAPFPFGNPINNNTMIMRSYI 194
      |||
QY 181 DGGLRVDTFRDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHSVRQHEIAGOI 240
      |||
DB 195 DGGLRVDTFRDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHSVRQHEIAGOI 254
      |||
QY 241 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
      |||
DB 255 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 314
      |||
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
      |||
DB 315 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 362
      |||

```

```

RESULT 4
QY3PS0 ID QY3PS0 PRELIMINARY; PRT; 346 AA.
AC QY3PS0;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Glycosyl transferase Lgta.
GN LGTA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PID2;
RA Tong Y., Arking D., Reinhold V., Stein D.C.;
  "Characterization of lipooligosaccharide structures found in Neisseria
  gonorrhoeae PID2.";

```


RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF313394; AAK70338.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Transference.
 SO SEQUENCE 346 AA; 39727 MW; 37C680CD23B149B4 CRC64;

Query Match 97.3%; Score 1773; DB 2; Length 346;
 Best Local Similarity 97.4%; Pred. No. 3e-144;
 Matches 339; Conservative 5; Mismatches 2; Indels 2; Gaps 1;

QY 1 MNPVSVLICAVNVEKYFQSLAAVYVNTWRNLDILYDDGSTDGLAIKPFQKRSRI 60
 DB 1 MNPVSVLICAVNVEKYFQSLAAVYVNTWRNLDILYDDGSTDGLAIKPFQKRSRI 60
 QY 61 KTLAQAONSGILPSLNLIGDELAKSGGGG--YIARTDADDDIAAPMIEKIVGEMEKDRSI 120
 DB 61 KTLAQAONSGILPSLNLIGDELAKSGGGG--YIARTDADDDIAAPMIEKIVGEMEKDRSI 118
 QY 121 IAMGAMLEYLESEKDGNNRLARRHKGKTIWKPTRHEDIAAEPFGNPINNTMIMRSYI 180
 DB 121 IAMGAMLEYLESEKDGNNRLARRHKGKTIWKPTRHEDIAAEPFGNPINNTMIMRSYI 178
 QY 119 IAMGAMLEYLESEKDGNNRLARRHKGKTIWKPTRHEDIAAEPFGNPINNTMIMRSYI 178
 DB 119 IAMGAMLEYLESEKDGNNRLARRHKGKTIWKPTRHEDIAAEPFGNPINNTMIMRSYI 178
 QY 181 DGLRYDTERDMAEDYQFVYDVSKGLRLAYPEALVKYRLHANOVSSKHSVQHEIAQSI 240
 DB 179 DGLRYDTERDMAEDYQFVYDVSKGLRLAYPEALVKYRLHANOVSSKHSVQHEIAQSI 238
 QY 241 OKTARNDFLOSGMFKTRPDSLEYROTAKAAYELPEKDLPEEDEFERARRFLYOCFKRTDP 300
 DB 239 OKTARNDFLOSGMFKTRPDSLEYROTAKAAYELPEKDLPEEDEFERARRFLYOCFKRTDP 298
 QY 301 PSGAWLDFEADGMRRLFTLRQYFGILYRLIKNRQARSAGKDEOI 348
 DB 299 PSGAWLDFEADGMRRLFTLRQYFGILYRLIKNRQARSAGKDEOI 346

RESULT 5

QYJW66 PRELIMINARY; PRT; 346 AA.
 AC Q9JW66;
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE Lacto-N-neotetraose blosynthesis glycosyl transferase.
 GN LGTA OR NMA0524.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxId=65699;
 RN NCBI_TaxId=65699;
 RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parshall J., Achman M., James K.D., Bentley S.D., Churcher C.,
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
 Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
 Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 Whitehead S., Spratt B.G., Barrell B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491".
 RL Nature 404:502-506(2000).
 DR EMBL: AL162753; CAB83816.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Complete proteome.
 SO SEQUENCE 346 AA; 40158 MW; 95C89CB1A0F1FFA9 CRC64;

Query Match 93.9%; Score 1712; DB 16; Length 346;
 Best Local Similarity 94.3%; Pred. No. 5.3e-139;
 Matches 328; Conservative 10; Mismatches 8; Indels 2; Gaps 1;

QY 1 MNPVSVLICAVNVEKYFQSLAAVYVNTWRNLDILYDDGSTDGLAIKPFQKRSRI 60
 DB 1 MNPVSVLICAVNVEKYFQSLAAVYVNTWRNLDILYDDGSTDGLAIKPFQKRSRI 60

DB 1 MNPVSVLICAVNVEKYFQSLATVYVNTWRNLEILYDDGSTDGLAIKPFQKRSRI 60
 QY 61 KTLAQAONSGILPSLNLIGDELAKSGGGGEYIARTDADDDISPGEIYVEMEKDRSI 120
 DB 61 KTLAQAONSGILPSLNLIGDELAKS--GMGEYIARTDADDDIAAPMIEKIVGEMEKDRSI 118
 QY 121 IAMGAMLEYLESEKDGNNRLARRHKGKTIWKPTRHEDIAAEPFGNPINNTMIMRSYI 180
 DB 119 IAMGAMLEYLESEKDGNNRLARRHKGKTIWKPTRHEDIAAEPFGNPINNTMIMRSYI 178
 QY 181 DGLRYDTERDMAEDYQFVYDVSKGLRLAYPEALVKYRLHANOVSSKHSVQHEIAQSI 240
 DB 179 DGLRYDTERDMAEDYQFVYDVSKGLRLAYPEALVKYRLHANOVSSKHSVQHEIAQSI 238
 QY 241 OKTARNDFLOSGMFKTRPDSLEYROTAKAAYELPEKDLPEEDEFERARRFLYOCFKRTDP 300
 DB 239 OKTARNDFLOSGMFKTRPDSLEYROTAKAAYELPEKDLPEEDEFERARRFLYOCFKRTDP 298
 QY 301 PSGAWLDFEADGMRRLFTLRQYFGILYRLIKNRQARSAGKDEOI 348
 DB 299 PSGAWLDFEADGMRRLFTLRQYFGILYRLIKNRQARSAGKDEOI 346

RESULT 6

QYJ8F0 PRELIMINARY; PRT; 322 AA.
 AC Q9J8F0;
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE Lgta.
 GN LGTA.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxId=485;
 RN NCBI_TaxId=485;
 RP SEQUENCE FROM N.A.
 RC STRAIN=F4899;
 RA Balhazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
 RT "Mutations in the lgt operon influence serum-resistance in
 gonococci".
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF208056; AAF25876.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 SO SEQUENCE 322 AA; 37109 MW; 88FECA5AD41DC2B7 CRC64;

Query Match 90.9%; Score 1657.5; DB 2; Length 322;
 Best Local Similarity 97.8%; Pred. No. 2.3e-134;
 Matches 316; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 26 VNOTWRNLDILYDDGSTDGLAIKPFQKRSRIKTLAQAONSGILPSLNLIGDELAKS 85
 DB 1 MNPVSVLICAVNVEKYFQSLATVYVNTWRNLEILYDDGSTDGLAIKPFQKRSRI 60
 QY 86 GGGGEYIARTDADDDIASPGWIEKIVGEMEKDRSI IAMGAMLEYLESEKDGNNRLARRHKGK 145
 DB 61 -GGGGEYIARTDADDDIASPGWIEKIVGEMEKDRSI IAMGAMLEYLESEKDGNNRLARRHKGK 119
 QY 146 GKIWKPTRHEDIAAEPFGNPINNTMIMRSYIDGRLRYDTERDMAEDYQFVYDVSKL 205
 DB 120 GAIWPKPTRHEDIAAEPFGNPINNTMIMRSYIDGRLRYDTERDMAEDYQFVYDVSKL 179
 QY 206 GRLAYPEALVKYRLHANOVSSKHSVQHEIAQSI OKTARNDFLOSGMFKTRPDSLEYRO 265
 DB 180 GRLAYPEALVKYRLHANOVSSKHSVQHEIAQSI OKTARNDFLOSGMFKTRPDSLEYRO 239
 QY 266 TKAAYELPEKDLPEEDEFERARRFLYOCFKRTDP PSGAWLDFEADGMRRLFTLRQYFG 325
 DB 240 TKAAYELPEKDLPEEDEFERARRFLYOCFKRTDP PSGAWLDFEADGMRRLFTLRQYFG 299
 QY 326 ILYRLIKNRQARSAGKDEOI 348

DB 300 ILRLIKNRQARSAGKEOEI 322

RESULT 7

ID 0918E9 PRELIMINARY; PRT; 321 AA.

AC 0918E9;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)

GN LGTA.

OS Neisseria gonorrhoeae.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=485;

RP [1]

RC SEQUENCE FROM N.A.

RC STRAIN=SS2;

RA Balhazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;

RT "Mutations in the lgt operon influence serum-resistance in RT gonococci";

RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF208059; AAF2877.1; -

DR InterPro: IPR001173; Glycos_transf.2.

DR Pfam: PF00535; Glycos_transf.2.1.

SQ SEQUENCE 321 AA; 37052 MW; 6ACADA9A3CB738FA CRC64;

Query Match 90.6%; Score 1651; DB 2; Length 321;

Best Local Similarity 97.5%; Pred. No. 8.4e-134;

Matches 315; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 26 VNQTWRNLDILIVDGSIDGTLAIANDFOKRDRIKILQAQNSGILPSINIGDELAKS 85

DB 1 MNQTWRNLDILIVDGSIDGTLAIANDFOKRDRIKILQAQNSGILPSINIGDELAKS 60

QY 86 GGGGGEYIARTDADDIASGWIETKIVGEMEKDRSIAMGAWLEVESEKDNRLAHNRH 145

DB 61 --GGGGEYIARTDADDIASGWIETKIVGEMEKDRSIAMGAWLEVESEKDNRLAHNRH 118

QY 146 GKIKKPTREHEDIAAFPPGPNPHNNTMTIMRSVIDGRLYDTERMAWDOYQWYVSKL 205

DB 119 GAIWKPTREHEDIAAFPPGPNPHNNTMTIMRSVIDGRLYDTERMAWDOYQWYVSKL 178

QY 206 GRLAYYPEALVYRLHANOVSSKHSVROHEIAQIQKTARNDFLOSGMGEFTREDSLEYRQ 265

DB 179 GRLAYYPEALVYRLHANOVSSKHSVROHEIAQIQKTARNDFLOSGMGEFTREDSLEYRQ 238

QY 266 TFAAAEELPEKDLPEDEFRARRFLYQCFKRTDTPPSGAWLDFPAAGRRRLFTLQYRFG 325

DB 239 TFAAAEELPEKDLPEDEFRARRFLYQCFKRTDTPPSGAWLDFPAAGRRRLFTLQYRFG 298

QY 326 ILRLIKNRQARSAGKEOEI 348

DB 299 ILRLIKNRQARSAGKEOEI 321

RESULT 8

Q9JX06 PRELIMINARY; PRT; 349 AA.

AC 09JX06;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)

DE Lacto-N-neotetraose biosynthesis glycosyl transferase LgtA.

GN NMB1929.

OS Neisseria meningitidis (serogroup B).

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=491;

RP [1]

RC SEQUENCE FROM N.A.

RC STRAIN=MC58 / SEROGROUP B;

RX MEDLINE=20175755; PubMed=10710307;

RA Tetelin B., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,

RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,

RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,

RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,

RA Mason T., Ciecko A., Parksey D.S., Blair E., Clitstone H., Clark E.B.,

RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vanachavan J.,

RA Gill J., Scariato V., Maignani V., Pizzo M., Grandi G., Sun L.,

RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;

RT "Complete genome sequence of Neisseria meningitidis serogroup B strain RT MC58";

RL Science 287:1809-1815(2000).

DR EMBL: AE002541; AAF42258.1; -

DR TIGR: NMB1929; -

DR InterPro: IPR001173; Glycos_transf.2.

DR Pfam: PF00535; Glycos_transf.2.1.

KW Transferase; Complete proteome.

SQ SEQUENCE 349 AA; 40618 MW; 4B9E245C34847DC3 CRC64;

Query Match 88.9%; Score 1621.5; DB 16; Length 349;

Best Local Similarity 92.5%; Pred. No. 3.3e-131;

Matches 309; Conservative 14; Mismatches 10; Indels 1; Gaps 1;

QY 1 MQLVSVLICATNVEKFFQSLAAVYNQTRNLDILIVDGSIDGTLAIANDFOKRDRI 60

DB 17 LQPLVSVLICATNVEKFFQSLAAVYNQTRNLDILIVDGSIDGTLAIANDFOKRDRI 76

QY 61 KITQAQNSGILPSINIGDELAKSGGGGEYIARTDADDIASGWIETKIVGEMEKDRSI 120

DB 77 RIIAQNRNSGILPSINIGDELAKS-GGGGEYIARTDADDIASGWIETKIVGEMEKDRSI 135

QY 121 IAWGAWLEVESEKDNRLARHHKGIWKKPTRHEDIAAFPPGPNPHNNTMTIMRSYI 180

DB 136 IAWGAWLEVESEKDNRLARHHKGIWKKPTRHEDIAAFPPGPNPHNNTMTIMRSYI 195

QY 181 DGLRLYDTERMAWDOYQWYVSKLGRILAYYPEALVYRLHANOVSSKHSVROHEIAQSI 240

DB 196 DGLRLYDTERMAWDOYQWYVSKLGRILAYYPEALVYRLHANOVSSKHSVROHEIAQSI 255

QY 241 QKTARNDFLOSGMGEFTREDSLEYRQTKAAYELPEKDLPEDEFRARRFLYQCFKRTDTP 300

DB 256 QKTARNDFLOSGMGEFTREDSLEYRQTKAAYELPEKDLPEDEFRARRFLYQCFKRTDTP 315

QY 301 PSGAWLDFPAAGRRRLFTLQYRFGILRLIKNR 334

DB 316 PSGAWLDFPAAGRRRLFTLQYRFGILRLIKNR 349

RESULT 9

Q51115 PRELIMINARY; PRT; 333 AA.

AC 051115;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)

GN Glycosyl transferase.

OS Neisseria meningitidis.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=487;

RP [1]

RC SEQUENCE FROM N.A.

RC STRAIN=MC58;

RX MEDLINE=96414473; PubMed=8817494;

RA Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;

RT "Molecular analysis of a locus for the biosynthesis and phase-variable RT expression of the lacto-N-neotetraose terminal lipopolysaccharide RT structure in Neisseria meningitidis";

RL Mol. Microbiol. 18:729-740(1995).

DR EMBL: U25839; AAC44084.1; -

DR InterPro: IPR001173; Glycos_transf.2.

DR Pfam: PF00535; Glycos_transf.2.1.

SQ SEQUENCE 333 AA; 38563 MW; 49D8F6CE375387BF CRC64;

Query Match 88.7%; Score 1617.5; DB 2; Length 333;

Best Local Similarity 92.5%; Pred. No. 6.8e-131;
Matches 309; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

QY 1 MOPVSVLICAYNEKFAOSLAIVNQTNRNLDILYVDSSTGTLIAAKPOKRSRI 60
Db 1 MOPVSVLICAYNEKFAOSLAIVNQTNRNLDILYVDSSTGTLIAARFEOBDR 60
QY 61 KIILAOQNSGLIPSLNIGLDELAKSGGGGGFYIARTDADIASPGWIEKIVGEMEKDRSI 120
Db 61 RIILQPNRNSGLIPSLNIGLDELAKS-GGGGGFYIARTDADIAADPWIEKIVGEMEKDRSI 119
QY 121 IAMGAMLEVLSEEDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMTIMRSVI 180
Db 120 IAMGAMLEVLSEEDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMTIMRSVI 179
QY 181 DGGRLYPTERMAEDYOFWYDVSKLGRLAYYPEALVYRLHANOVSSKSHVROHEIAGGI 240
Db 180 DGGRLYPTERMAEDYOFWYDVSKLGRLAYYPEALVYRLHANOVSSKSHVROHEIAGGI 239
QY 241 OKTARNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQCFKRTDTP 300
Db 240 OKTARNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQCFKRTDTP 299
QY 301 PSGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 334
Db 300 PAGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 333

RESULT 10
Q9EVD5 PRELIMINARY; PRT; 348 AA.
AC Q9EVD5;
DT 01-MAR-2001 (TReMBLrel, 16, Created)
DT 01-MAR-2001 (TReMBLrel, 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel, 21, Last annotation update)
DE N-acetylglucosamine transferase.
GN LGTA.
OS Neisseria subflava.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_Taxid=28449;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21142520; Pubmed=11208792;
RA Aking D., Tong Y., Stein D.C.;
RT "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae."
RL J. Bacteriol. 183:934-941(2001).
DR EMBL: AF240672; AAC09764.1; -.
DR HSSP: P39621; 10GQ.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KM Transferase.
SQ SEQUENCE 348 AA; 40676 MW; BC9E313E9BC0BF41 CRC64;

Query Match 88.3%; Score 1609; DB 2; Length 348;
Best Local Similarity 92.2%; Pred. No. 3.9e-130;
Matches 308; Conservative 13; Mismatches 11; Indels 2; Gaps 1;

QY 1 MOPVSVLICAYNEKFAOSLAIVNQTNRNLDILYVDSSTGTLIAAKPOKRSRI 60
Db 17 LQPLVSVLICAYNEKFAOSLAIVNQTNRNLDILYVDSSTGTLIAARFEOBDR 76
QY 61 KIILAOQNSGLIPSLNIGLDELAKSGGGGGFYIARTDADIASPGWIEKIVGEMEKDRSI 120
Db 77 RIILQPNRNSGLIPSLNIGLDELAKS-GGGGGFYIARTDADIAADPWIEKIVGEMEKDRSI 134
QY 121 IAMGAMLEVLSEEDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMTIMRSVI 180
Db 135 IAMGAMLEVLSEEDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMTIMRSVI 194
QY 181 DGGRLYPTERMAEDYOFWYDVSKLGRLAYYPEALVYRLHANOVSSKSHVROHEIAGGI 240
Db 195 DGGRLYPTERMAEDYOFWYDVSKLGRLAYYPEALVYRLHANOVSSKSHVROHEIAGGI 254

QY 241 OKTARNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQCFKRTDTP 300
Db 255 OKTARNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQCFKRTDTP 314

QY 301 PSGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 334
Db 315 PAGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 348

RESULT 11
Q93EK6 PRELIMINARY; PRT; 221 AA.
AC Q93EK6;
DT 01-DEC-2001 (TReMBLrel, 19, Created)
DT 01-DEC-2001 (TReMBLrel, 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel, 19, Last annotation update)
DE LGTA.
GN LGTA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_Taxid=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M978;
RX MEDLINE=21467954; Pubmed=11563844;
RA Zhu P., Klutch M.J., Tsai C.-M.;
RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide Expression in Two 18-Immunotype Strains of Neisseria meningitidis."
RL FEMS Microbiol. Lett. 203:173-177(2001).
DR EMBL: AF353193; AAL12840.1; -.
SQ SEQUENCE 221 AA; 26478 MW; D1F97099B1F6D55F CRC64;

Query Match 94.3%; Score 1118; DB 2; Length 221;
Best Local Similarity 94.6%; Pred. No. 3.5e-88;
Matches 209; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 114 MEKDRSIAMGAMLEVLSEEDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMT 173
Db 1 MEKDRSIAMGAMLEVLSEEDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMT 60
QY 174 IMRSYVDGRLYPTERMAEDYOFWYDVSKLGRLAYYPEALVYRLHANOVSSKSHVRO 233
Db 61 IMRSYVDGRLYPTERMAEDYOFWYDVSKLGRLAYYPEALVYRLHANOVSSKSHVRO 120
QY 234 HEINOIGIOKTARNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQCF 293
Db 121 HEINOIGIOKTARNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQCF 180
QY 294 FKRTDTPPSGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 334
Db 181 FKRTDTPPSGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 221

RESULT 12
Q50949 PRELIMINARY; PRT; 337 AA.
AC Q50949;
DT 01-NOV-1996 (TReMBLrel, 01, Created)
DT 01-NOV-1996 (TReMBLrel, 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel, 17, Last annotation update)
DE Glycosyl transferase.
GN LGTD.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_Taxid=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F62;
RX MEDLINE=95053752; Pubmed=7964493;
RA Gotschlich E.C.;
RT "Genetic locus for the biosynthesis of the variable portion of Neisseria gonorrhoeae lipooligosaccharide."
RL J. Exp. Med. 180:2181-2190(1994).

```
DR EMBL; U14554; AAA68012.1; -.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Transferase
SQ SEQUENCE 337 AA; 38393 MW; 5AE21AA75CAA732E CRC64;

Query Match 57.5%; Score 1049; DB 2; Length 337;
Best Local Similarity 64.5%; Pred. No. 5,6e-82;
Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

QY 1 MQLPVSLICAYNVEKYFAOSLAAYVNOTWRNDLIIYDDGSTDGTALAKDFQKRSRI 60
DB 1 MQLPVSLICAYNVEKYFAOSLAAYVNOTWRNDLIIYDDGSTDGTALAKDFQKRSRI 60
QY 61 KILAOAONSGILPSLNLGDELAKSGGGGEYIARTDADDIASPGMIKIVGEMKDRSI 120
DB 61 KILAOAONSGILPSLNLGDELAKSGGGGEYIARTDADDIASPGMIKIVGEMKDRSI 118
QY 121 IAMGAMLEVLSEEDGNRLARHHKGIWKKPTRHEDIAAEFFPGNPJHNTMIRRSVI 180
DB 119 IAMGAMLEVLSEENKSVLAIAIRNGALIMDKPTRHEDIVAVPEFGNPJHNTMIRRSVI 178
QY 181 DGLRPTDERMAEDYQWYDVYSKIGRLATYPEALVKRLANOVSSKHSVROHEIAQI 240
DB 179 DGLRPTDERMAEDYQWYDVYSKIGRLATYPEALVKRLANOVSSKHSVROHEIAQI 238
QY 241 OKTARNDLQSMGFKTRDSLEYROTAKAAAYELPEKDLPEEDFERARRELYOCFRTDTP 300
DB 239 KEIRIAGYKWAAGIAGVADCLNTYGLKSTAYALYKALKSGDIGCLRFLYEVFLSLEKY 298
QY 301 PSGAWLDFPADGRMRRLFTLRQYFGLYRLIK 332
DB 299 SLTDLDFLTDRVMKRLFAAPQYRKILKMLR 330

RESULT 13
Q9CLR9 PRELIMINARY; PRT; 337 AA.
AC Q9CLR9;
DT 01-JUN-2001 (TREMblrel. 17; Created)
DT 01-JUN-2001 (TREMblrel. 17; Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19; Last annotation update)
DE Hypothetical protein PM1140.
GN PM1140.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang O., Li L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Fm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006155; AAK03224.1; -.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 337 AA; 39267 MW; 8639BCFB5F700D84 CRC64;

Query Match 36.6%; Score 666.5; DB 16; Length 337;
Best Local Similarity 42.9%; Pred. No. 4,5e-49;
Matches 145; Conservative 59; Mismatches 113; Indels 21; Gaps 5;

QY 2 QPLVSLICAYNVEKYFAOSLAAYVNOTWRNDLIIYDDGSTDGTALAKDFQKRSRI 61
DB 6 QPLVSLICAYNVEKYFAOSLAAYVNOTWRNDLIIYDDGSTDGTALAKDFQKRSRI 65
QY 62 ILAQAONSGILPSLNLGDELAKSGGGGEYIARTDADDIASPGMIKIVGEMKDRSI 121
DB 66 ILAQAONSGILPSLNLGDELAKSGGGGEYIARTDADDIASPGMIKIVGEMKDRSI 119
```

```
QY 122 AMGAMLEVLSEEDGNRLARHHKGIWKKPTRHEDIAAEFFPGNPJHNTMIRRSVI 180
DB 120 AMGAYLFTLSDDGNSNLANYEHGDEWRNPLSHREIVEAMLFRNPJHNTMIRRSVI 179
QY 181 DGLRPTDERMAEDYQWYDVYSKIGRLATYPEALVKRLANOVSSKHSVROHEIAQI 240
DB 180 EHGLRFPDPAHYQHTDEYQFMLEVSRIGELANYPESILVYRLHNTQTSJLHNYONIMARKI 239
QY 241 OKTARNDLQSMGFKTRDSLEYROTAKAAAYELPEKDLPEEDFERARRELYOCF 294
DB 240 KRRANITLQDLGVYHRLGEDIFHDITITQAEIASLSLONCI-----IKRLIYDCY 292
QY 295 KRTDTPSGAWLDFPADGRMRRLFTLRQYFGLYRLIK 332
DB 293 LSLVDMKLNILYFLRD-KNSYEFKKOKIKIRRIIR 329

RESULT 14
P74947 PRELIMINARY; PRT; 337 AA.
P74947;
AC P74947;
DT 01-FEB-1997 (TREMblrel. 02; Created)
DT 01-MAY-1999 (TREMblrel. 10; Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21; Last annotation update)
DE ORF139-19 protein (putative glycosyltransferase A).
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MO45;
RA Yamasaki S., Shimizu T., Hoshino K., Ho S., Shimeda T., Nair G.B.,
RA Takeda Y.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 157-337 FROM N.A.
RC STRAIN=AI 1838;
RX MEDLINE=97094999; PubMed=8940420;
RA Falkind S., Stark M., Albert M.J., Uhlen M., Lundeberg J.,
RA Weintraub A.;
RT "Cloning and sequence of a region of Vibrio cholerae O139 Bengal and
RT its use in PCR-based detection.";
RL J. Clin. Microbiol. 34:2904-2908(1996).
DR EMBL; AB012956; BAA33603.1; -.
DR EMBL; U72485; AAC44837.1; -.
DR HSSP; P39621; 10G0.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Transferase.
SQ SEQUENCE 337 AA; 39164 MW; 721D748FCB8C9798 CRC64;

Query Match 22.6%; Score 412.5; DB 2; Length 337;
Best Local Similarity 33.0%; Pred. No. 3,2e-27;
Matches 104; Conservative 55; Mismatches 113; Indels 43; Gaps 9;

QY 3 PLVSVLICAYNVEKYFAOSLAAYVNOTWRNDLIIYDDGSTDGTALAKDFQKRSRI 62
DB 4 PRISVMSVMEKYEKLEGAIDSIILKQTFSDPEFIINDGSDTKLEIKSYWKKDRI-V 62
QY 63 LAQAONSGILPSLNLGDELAKSGGGGEYIARTDADDIASPGMIKIVGEMKDRSI 122
DB 63 LVSRENKGLIVSLNGLD-LAK----GQYIARMADADISIKSREKQIEFLDSNPDIGV 116
QY 123 MGAMLEVLSEEDGNRLARHHKGIWKKPTRHEDIAAEFFPGNPJHNTMIRRSVID 182
DB 117 CGTWVEVGE-----NKSKKMKAPTOPDPRLAKLFLVSFPAIPTWIRKRVIDK 166
QY 183 GLRPTDERMAEDYQWYDVYSKIGRLATYPEALVKRLANOVSSKHSVROHEIA 237
DB 167 YKINDLQYKMAEDYKFWVDSESKYTLFNVPEILRLRYHQDESIRVADNKNENKRFELI 226
QY 238 OGIO-----KTARNDLQSMGFK-----TFPDSLEYROTAKAAAYELPEKDL 279
```

Db 227 SKIDNEVLTSIGIVLTNEGAKNHFTLSLNERITNNVYDCMIRAHLLKISSQIESSQFD 286

QY 280 EEDFER---ARFLY 291

Db 287 SSATERLMLTKRYFLY 301

RESULT	15
0343234	
ID	0343234
AC	0343234;
DT	01-JAN-1998 (TREMblrel. 05, Created)
DT	01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT	01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE	Sugar transferase.
OS	Vibrio cholerae.
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX	NCBI_TaxID=666;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=AT-1837;
RX	MEDLINE=97252505; PubMed=9098074;
RT	Strecher U.H., Paraisivam G., Dredge B.K., Manning P.A.;
RT	"Novel Vibrio cholerae O139 genes involved in hippopolysaccharide biosynthesis.";
RL	J. Bacteriol. 179:2740-2747(1997).
DR	EMBL; Y07786; CAA69119.1; -.
DR	HSSP; P39621; 10Q.
DR	InterPro; IPR001173; Glycos_transf_2.
DR	Pfam; PF00535; Glycos_transf_2, 1.
DR	transferase.
Q0	SEQUENCE 337 AA: 39236 MW: 5630851088B135464 CRC64;

Query Match	22.5%	Score 410;	DB 2;	Length 337;
Best Local Similarity	37.9%	Pred. No. 5.3e-27;		
Matches 97;	Conservative 45;	Mismatches 88;	Indels 26;	Gaps 7;

```

QY 3 PLYSVLCAVNVKVEYFQOSLAAVYNWTRLDLIYDDGSTOCTLAADFOKDSRIRI 62
   | : ||| : ||| : : : ||| : : : ||| ||| | : ||| : ||| |
Db 4 PRISVWASVYNGEYUJGEALDLSILKQTFSPFEPIITINDGSTDTLELISYMKKDRIV 62
   | : ||| : ||| : : : ||| : : : ||| ||| | : ||| : ||| |
QY 63 LAQONSGLIPSLNIGLDELAKSGGGGGEYIARTDADDIASPGWMEIKYGEEMEKDSIIA 122
   | : ||| : ||| : : : ||| : : : ||| ||| | : ||| : ||| |
Db 63 LVSRENKGLIYSLNEGID-LAK-----GQYIARDADDISIKSRFEQJEFELSDNDICV 116
   | : ||| : ||| : : : ||| : : : ||| ||| | : ||| : ||| |
QY 123 MGARLEVLSEKQOSNRLLARRHHKKGITWKKRTREDLIAPFRFENPILHNMTMRBSVIOG 182
   | : ||| : ||| : : : ||| : : : ||| ||| | : ||| : ||| |
Db 117 CGIWEVVEFG-----NIKSKMKMTOPDPLAKLTFSVPRANPRTVIAKKVYIDK 166
   | : ||| : ||| : : : ||| : : : ||| ||| | : ||| : ||| |
QY 183 -GLRYEDRWMAEDYQFWYDVSKIGRLAAYREALVYKRLANOVS-----SKHSVROHEIA 237
   | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 167 YKINVDYGDMAEDYKRWVDFSKYTLFSSNPELLRLRYRQESISNADNKEKKEFEIIT 226
   | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 238 QGIQKTARNDFLQSMG 253
   | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 227 SKIQ-----NEVLTSVG 238
   | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

```

Search completed: December 2, 2002, 12:02:15
Job time : 31.8926 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 31.7609 Seconds
(without alignments)
1283.801 Million cell updates/sec

Title: US-10-007-267-4
Perfect score: 1657
Sequence: 1 MDIVFAADNDNAAAYLCVAAK.....KCMIGRWKKLSARFLRKITY 306

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1657	100.0	306	17 AAR91313	N. gonorrhoeae gly
2	1657	100.0	306	18 AAM06578	Lipo-oligosacchari
3	583	33.2	106	22 AAB72456	Partial GlyT. Nei
4	572	34.5	302	21 AAB44566	Virulence gene pro
5	568	34.3	302	21 AAB44538	Human 67210 glycos
6	330	19.9	334	23 AAU9917	Partial GlyT. Hae
7	280	16.9	106	22 AAB72455	Streptococcus pneu
8	233	14.1	399	21 AAY81561	Neisseria meningit
9	205	12.4	50	20 AAM89327	Streptococcus poly
10	203.5	12.3	401	23 ABB28548	Streptococcus poly

11	197.5	11.9	332	21 AAG15767
12	197.5	11.9	346	21 AAG15766
13	197	11.9	313	21 AAY91285
14	197	11.9	398	23 ABB28550
15	196	11.8	312	23 ABB27411
16	190.5	11.5	332	21 AAG45491
17	190.5	11.5	346	21 AAG45490
18	190.5	11.5	346	23 ABB92808
19	186	11.2	361	23 ABB90816
20	184	11.1	272	23 ABB27661
21	184	11.1	392	21 AAG28027
22	184	11.1	409	21 AAG28026
23	184	11.1	441	21 AAG28025
24	178.5	10.8	401	23 ABB27413
25	176.5	10.7	674	19 AAM61238
26	176.5	10.7	674	23 ABB54656
27	175.5	10.6	413	23 ABB27412
28	171.5	10.4	357	21 AAG13328
29	171.5	10.4	357	23 ABB92767
30	171.5	10.4	361	21 AAG13327
31	169.5	10.2	351	23 ABB92218
32	165	10.0	323	21 AAG10403
33	165	10.0	364	21 AAG10402
34	165	10.0	365	21 AAG10401
35	162.5	9.8	341	23 ABB92615
36	157	9.5	305	21 AAG55049
37	157	9.5	372	19 AAM98360
38	157	9.5	372	19 AAB60318
39	157	9.5	384	21 AAG30296
40	157	9.5	390	23 AAG30295
41	157	9.5	390	23 ABB91529
42	157	9.5	382	21 AAG13970
43	156.5	9.4	345	21 AAG13969
44	156.5	9.4	385	21 AAG13968
45	156.5	9.4	365	21 AAG13968

ALIGNMENTS

RESULT 1	
AA91313	AA91313 standard; Protein; 306 AA.
ID	AA91313;
AC	AA91313;
DT	09-JUL-1996 (first entry)
XX	N. gonorrhoeae glycosyltransferase LgtC.
DE	
XX	
KW	Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus; vaccine.
OS	Neisseria gonorrhoeae strain F62.
XX	
PN	W09610086-A1.
XX	
PD	04-APR-1996.
XX	
PF	25-SEP-1995; 95WO-US12317.
XX	
PR	26-SEP-1994; 94US-0312387.
XX	
PA	(UTRG) UNIV ROCKEFELLER.
XX	
PI	Gotschlich EC;
DR	WPI: 1996-200924/20.
XX	
DR	N-PSDB; AAT14061.
XX	
PT	Nucleic acids encoding glycosyl transferase(s) - used in the diagnosis of infection with Neisseria and for the biosynthesis of oligo:saccharide(s)

XX Claim 10; Fig 2d; 81pp; English.

XX

PS

XX

CC 5 Glycosyltransferases (AA93131-15) are products of the 1gt locus

CC (AA914061) of *Neisseria gonorrhoeae* strain F62. Glycosyltransferase

CC LgtC (AA93133) can be obt'd. by expression of the lgtC coding

CC sequence in recombinant host cells. A method for adding Gal

CC alpha1-4 to Gal comprises contacting a reaction mixture contg.

CC activated Gal to an acceptor moiety comprising a Gal residue in the

CC presence of LgtC. Oligosaccharides can be produced that, when

CC attached to non-toxic lipids, are useful for *Neisseria* vaccine prepn.

CC Blood group core oligosaccharides, and mimics of lacto-N-neotetraose,

CC gangliosides and saccharide portions of glyboglcolipids can also be

CC produced using the enzymes.

XX

XX

Sequence 306 AA;

Query Match	100.0%	Score 1657	DB 17	Length 306
Best Local Similarity	100.0%	Pred. No. 4,8e-169		
Matches 306	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	1	MDIYFAADNDNAAALCYAANKVEAHBDTEIRFVLDAGTSEENRAVAANLGGGNIR	60
Db	1	MDIYFAADNDNAAALCYAANKVEAHBDTEIRFVLDAGTSEENRAVAANLGGGNIR	60
QY	61	IDVNPDEBFAFPLMRIHISTTYARLKLGEYIADCDKVLVLTDTVLRDGLKPLMDTDG	120
Db	61	IDVNPDEBFAFPLMRIHISTTYARLKLGEYIADCDKVLVLTDTVLRDGLKPLMDTDG	120
QY	121	GNMWGACIDLEVEENOEGKOKIGADEEYFNAGVLLINKKVRRHDIFFMGCSEWDEYOK	180
Db	121	GNMWGACIDLEVEENOEGKOKIGADEEYFNAGVLLINKKVRRHDIFFMGCSEWDEYOK	180
QY	181	DVMOYODDILINGLFKGGVCYANSRFNEMPTNAAFMANGFASRHTDPLYLDRINTAMPYA	240
Db	181	DVMOYODDILINGLFKGGVCYANSRFNEMPTNAAFMANGFASRHTDPLYLDRINTAMPYA	240
QY	241	VSHCGSAAKPMWHRDCTYWGAEERFTELGSLLTYPEEMRGKFLAVPPIKCMLOBRKRLTSAR	300
Db	241	VSHCGSAAKPMWHRDCTYWGAEERFTELGSLLTYPEEMRGKFLAVPPIKCMLOBRKRLTSAR	300
QY	301	FLRKITY 306	
Db	301	FLRKITY 306	

RESULT 2	
AAW06578	
ID	AAW06578 standard; Protein; 306 AA.

AC	AAW06578;	
XX		
DT	21-MAR-1997	(first entry)
XX		
DE	Lipo-oligosaccharide gene-encoded protein.	
XX		
KW	Polyglycosyltransferase; N-acetylglucosami	
KW	N-acetylglactosaminyl transferase; lipo-c	
XX		
OS	Neisseria gonorrhoeae ATCC 33084.	
XX		
FN	W096640971-11.	
XX		
PD	19-DEC-1996.	
XX		
PF	03-JUN-1996;	96MO-US08323.
XX		
PR	07-JUN-1995;	95US-0478140.
XX		
PA	(NEOS-) NEOSE TECHNOLOGIES INC.	
XX		
PI	Buczala SL, Johnson KF, Roth S;	
XX		

DR WPI; 1997-052351/05.
DR N-PSDB; AAT49230.

PT Transfer of at least 2 saccharide units using
PT poly:glycosyl:transferase - isolated from *N. gonorrhoeae*, catalyses
PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
PT single galactose moiety

PS Disclosure; Fig 2C-E; 38pp; English.

CC A lipooligosaccharide-encoding gene region (Aat49230) of *Neisseria*
CC gonorrhoeae ATCC 33084. Includes coding sequences for 5 proteins
CC (AAM06576-80), one of which (AAM06576) is a polyglycosyltransferase
CC that catalyses the addition of GlcNAc and GalNAc disaccharides to
CC a galactose moiety. The function of the other 4 proteins is not
CC stated in the specification.

AA	Sequence	306 AA;
SQ		

Query Match	100.0%	Score 1657;	DB 18;	Length 306;
Best Local Similarity	100.0%;	Pred. No. 4;	8e-165;	
Matches 306; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	MDIPEADDDNNYAALCYAANSVEAAHDDTEIRFHVLDAGISEENRAVAANLRGGGNIREF	60
Db	1	MDIPEAADDDNNYAALCYAANSVEAAHDDTEIRFHVLDAGISEENRAVAANLRGGGNIREF	60
QY	61	IDVNEPEDEPAGEPLNI RHISITTYARLKGEYIADCDKVLVLDTDVLRDGIKPLMTDILG	120
Db	61	IDVNEPEDEPAGEPLNI RHISITTYARLKGEYIADCDKVLVLDTDVLRDGIKPLMTDILG	120
QY	121	GNMWGACICDIDVEBOEGKKQIGMADDEYFPMNGVLLINKKWRHRHIDFPMSCSEWQYOK	180
QY	121	GNMWGACICDIDVEBOEGKKQIGMADDEYFPMNGVLLINKKWRHRHIDFPMSCSEWQYOK	180
Db	121	GNMWGACICDIDVEBOEGKKQIGMADDEYFPMNGVLLINKKWRHRHIDFPMSCSEWQYOK	180
QY	181	DVMOYODODILNGFLEKSGVCYANSRPFNFMPTNYAFMANGFASRHTDPLYLDRNTTAMPVA	240
Db	181	DVMOYODODILNGFLEKSGVCYANSRPFNFMPTNYAFMANGFASRHTDPLYLDRNTTAMPVA	240
QY	241	VSHHGSAKKPMWHRCTWAGARFELAGSLTTTVEEARGKLAVPPTCMLORMPKKTLSAR	3000
Db	241	VSHHGSAKKPMWHRCTWAGARFELAGSLTTTVEEARGKLAVPPTCMLORMPKKTLSAR	3000
QY	301	FLRKITY 306 	
Db	301	FLRKITY 306 	

RESULT 3
AAB72456
ID AAB72456 standard; Protein; 106 AA.

AC	AAB72456;	
XX		
DT	08-MAY-2001	(first entry)
XX		
DE	Partial GLYT.	
XX		
KM	UGGT; UDP-glucose:glycoprotein glucosyltransferase; enzyme: GLYT	
XX		
CS	Neisseria gonorrhoeae.	
XX		
FN	WO200112845-A1.	
XX		
PD	22-FEB-2001.	
XX		
PE	27-JUL-2000;	2000MO-CA00883.
XX		
FR	18-AUG-1999;	99US-0376330.
XX		
PA	(CANA) NAT RES COUNCIL CANADA.	
XX		
DT	Tessier DC, Dignard D, Bergeron JMM, Thomas DY;	

XX		WP1; 2001-218358/22.
DR		
PT	Determining the effect of a test sample on UDP-glucose:glycoprotein	
PT	glucosyltransferase (UGGT), useful for measuring UGGT activity,	
PT	comprises exposing an acceptor substrate for UGGT to a labeled donor in	
PT	the presence of UGGT -	
XX		
PS	Disclosure; Fig 10; 95pp; English.	
CC	The present invention relates to a method for determining the effect of a	
CC	test sample on UDP-glucose:glycoprotein glucosyltransferase (UGGT)	
CC	activity. The method comprises exposing an acceptor substrate for UGGT to	
CC	a labelled donor in the presence of the test sample and UGGT. The method	
CC	is useful for determining UGGT activity. In particular, the method is	
CC	useful in glucosyltransferase assay and kinetics measurement for	
CC	determining UGGT activity. UGGT is a soluble enzyme of the endoplasmic	
CC	reticulum which catalyses the addition of a glucose residue onto	
CC	asparagine-linked oligosaccharides, which are present on incorrectly	
CC	folded glycoproteins. The present sequence was used in a sequence	
CC	homology comparison with rat UGGT (see AAF60732 and AAB72436) which was	
CC	used in the method of the present invention.	
CC		
SQ	Sequence 106 AA;	
	Query Match 35.2%; Score 583; DB 22; Length 106;	
	Best Local Similarity 100.0%; Pred. No. 1.9e-54;	
	Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY	92 IADCKVLYLTDVLVRDGLKPLMTDLGNNVGACIDLFVEROESYKOKIGMAGEYYF 151 	
Db	1 IADCKVLYLTDVLVRDGLKPLMTDLGNNVGACIDLFVEROESYKOKIGMAGEYYF 60 	
QY	152 NAGVLLINLKRRRHDIFFKSCFEWEOYKDVAQODILNGLFKG 197 	
Db	61 NAGVLLINLKRRRHDIFFKSCFEWEOYKDVAQODILNGLFKG 106 	
RESULT 4		
AAB44566		
ID	AAB44566 standard; Protein; 302 AA.	
XX		
AC	AAB44566;	
DT	08-FEB-2001 (first entry)	
DE	Virulence gene protein #46.	
XX		
KW	Virulence gene; antibacterial; vaccine; bacterial infection;	
KW	septicemia; Bronchopneumonia; rhinitis; wound infection.	
XX		
OS	Pasteurella multocida.	
PN	MO200061724-A2.	
PD	19-OCT-2000.	
XX		
PF	06-APR-2000; 2000WO-US09218.	
XX		
PR	09-APR-1999; 99US-0128689.	
PR	10-SEP-1999; 99US-0153453.	
XX		
PA	(PHAA) PHARMACIA & UPJOHN INC.	
XX		
PI	Lowery DE, Fuller TE, Kennedy MJ;	
XX		
DR	WP1; 2000-647422/62.	
DR	N-PSDB; AAC79641.	
XX		
PT	Attenuated Pasteurellaceae bacteria comprising mutations in virulence	
PT	genes, useful as a live attenuated vaccine against bacterial infections	
XX		

PS Claim 39; Pages 260-261; 322pp; English.

xx The family Pasteurellaceae encompasses several pathogens that infect a
cc wide variety of animals. The present invention relates to virulence genes
cc from Pasteurellaceae. The present sequence is a protein encoded by one
cc such virulence gene. The virulence genes of the present invention may be
cc mutated in order to produce an inactive gene. The inactive virulence gene
cc may in turn be used to produce a vaccine, which is useful for treating
cc bacterial infections such as septicemias, bronchopneumonias, rhinitis and
cc wound infections.

xx Sequence 302 AA:
SO

Query Match 34.5%; Score 572; DB 21; Length 302;
Best Local Similarity 38.9%; Pred. No. 1,4e-52;
Matches 121; Conservative 61; Mismatches 115; Indels 14; Gaps 7

OY 1 MDIVPADDDNNYAAYLCVAAKSYEAHAHPDEIRFPHVLDAIGISEENRAAVANLARG-GGNIR 59
Db 1 NMLIEVSDDVYAKHLVAALKSI-INHNNEGISFTYIDLSIKDENKNINDIYSSTSEVN 59
OY 60 FIDVPNEDFEAGPELNIRHISITTYARLKLGEXIAD-CDKVLYLDTDVYLVDGLKPLMDTD 118
Db 60 FLAVNEKEPESPFGDISYSLATFYARKAEYLPDLNKITIYLDVDVLFNSLEMLMNVND 119
OY 119 LGGNNVGACIDLFPVRQES-YKOKIGMADEGYFNAGULLNLKKRRRDIFKMSCEWE 177
Db 120 VNNFLTACCYDSFIENEKSEHKRSISMSPDKXEYFNAGVALFLNDEMRKMDFSRALDILA 179
OY 178 QYKDMMQODODILNGLFEGYCVANSRPNEMPTNAPAFANGFASHHTDPYLDRIINTAM 237
Db 180 MPDNMTIYODODIILNFLENKKVCYLDCKRPENPNOLERIKOYHGKGLSNLSLEK--TTM 237
OY 238 PVAVSHYCGSAKPWHRDCTVWGAEFFTELAGSLTTYPREMG--KLAVPTKCMLORMWR 295
Db 238 PVVISHYCGEPKAMADCKHFNVIFYQKI LAET-----RGDKERNVLSITYLKALIR 291
OY 296 KLSARFLRKII 306
Db 292 RIRYKFKYQYV 302

RESULT 5
AAB44538
ID AAB44538 standard; Protein; 302 AA.
XX
XX AAB44538;
XX AC
XX DT 08-FEB-2001 (first entry)
XX XX
DE Virulence gene protein #18.
XX XX
KW Virulence gene; antibacterial; vaccine; bacterial infection;
KW septicemia; bronchopneumonia; rhinitis; wound infection.
XX OS
OS Pasteurella multocida.
XX PN
PN WO200061724-A2.
PD 19-OCT-2000.
PF 06-APR-2000; 2000WO-US09218.
PR 09-APR-1999; 99US-0128689.
PR 10-SEP-1999; 99US-0153453.
XX PA (PHAA) PHARMACIA & UPJOHN INC.
XX PI Lowery DE, Fuller TE, Kennedy MJ;
DR WPI: 2000-647422/62.
DR N-PSTDB; AAC79598.
XX

PT Attenuated Pasteurellaceae bacteria comprising mutations in virulence
PT genes, useful as a live attenuated vaccine against bacterial infections
PT -
XX
PS Disclosure; Pages 149-150; 322pp; English.

PS Disclosure; Pages 149-150; 322pp; English..

The family Pasteurellaceae encompasses several pathogens that infect a wide variety of animals. The present invention relates to virulence genes from Pasteurellaceae. The present sequence is a protein encoded by one such virulence gene. The virulence genes of the present invention may be mutated in order to produce an inactive gene. The inactive virulence gene may in turn be used to produce a vaccine, which is useful for treating bacterial infections such as septicemias, bronchopneumonias, rhinitis and wound infections.

... SQ Sequence 302 AA;

Query Match	34.38	Score 568	DB 21	Length 302
Best Local Similarly	38.68	Pred. No. 3.7e-52		
Matches 120; Conservative	61	Mismatches 116	Indels 14	Gaps 7

Qy	1	MDIYEADDDNNAALCYAAKSVENAAHDDTEIRFVLDAGLSEENRAVAANLIR-GENIR	59
Dd	1	MNILEFSDDYAKRLVAAIKSI-INHNEMKGSFIEDPLGIDKEXKRNINIVSSYSEVN	59
Qy	60	FIDVNPEDFAGPLNIRHISTTYTARLKGLEYIAD-CDKVLVDTDVLRDGLKPLMDTD	118
Dd	60	FIAVNEKEFESEFPPOISISTATYARLKAEEYLEDPDNINKITTYEDVLPVNSIEMLNVD	119
Qy	119	LGGMWVGACIDLFERBOG-VKQIGAMADGEYYPNAGVLLINKKKMRHIDFKKSCWE	177
Dd	120	VNNELTACIYCSFLENKSEHKRISMSDKREYTYNAGVYMLFNDEWKKMDYFSALDILA	179
Qy	178	QYKDVMQYQDODILNGLFKGVCYANSRFPNPTNYAFMANGFASRHTDPLDYDRTNIA	237
Dd	180	MYPMQMITQDDDIINIIFRNKVCYCLDRFENPMQULERIXQYHKKGSNNLSLEK--TTM	237
Qy	238	PVAVSHVGSAAKPMWHRDCTVWGAARFETVAGSLTVEEMRG--KLAVPPTKMLQWRK	295
Dd	238	PVVISHTYGGPKAMHADCKHNNYFYQKILAXS-----RGXDKERYLSIKTYLKLALIR	291
Qy	296	KLASAFLEKRIY 306	
Dd	292	RIRYKFKQYV 302	

RESULT 6
AAU99917
ID AAU99917 standard; Protein; 334 AA.

DT	07-OCT-2002 (first entry)
XX	
DE	Human 67210 glycosyl transferase domain consensus sequence

KW 67210; glucosyl transferase; infection;
KW haematopoietic disorder; blood clotting disorder; cancer;
KW autoimmune disorder; leukaemia; immunological disorder;
KW cardiovascular disorder; neurological disorder; cellular proliferation;
KW red blood cell disorder; viral disease; neurological disorder.

Synthetic.

PN W0200240656-A2.

22 23-MAY-2002

14-NOV-2001: 2001WQ-TTSA5291

XX
14-NOV-2000 2000TC-249231D

PR 14-NOV-2000; 2000US-248362P.

PR	30-NOV-2000;	2000US-250077P
PR	30-NOV-2000;	2000US-250176P
PR	30-NOV-2000;	2000US-250337P
XX	(MILL-) MILLENNIUM PHARM INC	
PA		
XX		
PI	Meyers RE, Curtis RAD, GL	
XX		
DR	WPI; 2002-508325/54.	

PA (MILL-) MILLENNIUM PHARM INC.

PI Meyers RE, Curtis RAJ, Glucksmann MA;

DR WPI; 2002-508325/54.

PT Isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
PT or 84334 polypeptides, useful as reagents or targets for treating or
PT diagnosing pain or metabolic, liver, kidney, or cardiovascular
PT disorders -

PS Disclosure; Flg 6; 298pp; English.

CC This invention relates to the DNA and protein sequences of novel
CC Isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
CC or 84234 proteins. The method of the invention is useful for treating a
CC disorder characterised by aberrant activity of 47476, 67210, 49875,
CC 46842, 33201, 83378, 84233, 64708, 85041 or 84234-expressing cell in a
CC subject. The protein molecules can act as novel diagnostic targets and
CC therapeutic agents for controlling aberrant or deficient signal
CC transduction resulting, in e.g., haematopoietic disorders, including
CC blood clotting disorders, autoimmune disorders, or disorders related to
CC an inability to clear infections (e.g., viral or bacterial infections),
CC as well as disorders related to abnormal cellular proliferation or
CC differentiation, e.g., leukaemia. They may also be used to control
CC disorders of metabolic imbalance (e.g., disorders of lipopolysaccharide
CC biosynthesis or glycogen synthesis) immunological disorders,
CC cardiovascular disorders, neurological disorders, or cellular
CC proliferation and/or differentiation disorders, e.g., cancer, cell
CC motility and adhesion disorders differentiative disorders, red blood
CC cell disorders, viral diseases, neurological disorders (e.g., brain
CC disorders), pain or metabolic disorders, liver disorders, kidney
CC disorders, disorders of the small intestine, disorders of metal ion
CC imbalance, protein trafficking disorders and disorders associated with
CC bone metabolism. The sequences of the invention are also useful for
CC screening assays, predictive medicine (e.g., diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenetics);
CC and methods of treatment (e.g., therapeutic and prophylactic). The
CC present sequence represents a predicted consensus sequence motif
CC found in the protein of the invention.

... Sequence 334 AA:

Query Match	19.98;	Score 330;	DB 23;	Length 334;
Best Local Similarity	34.98;	Pred. No. 1.5e-26;		
Matches 102; Conservative	33;	Mismatches 87;	Indels 70;	Gaps 15

```

QY 28 DTEIHFVLDAGISENSEAAVAANLRGGN-----IRFDVAPED- F 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17 ENNVFHLITDGFCKRSLLYSQ-----GDNNKLLTETLAPPEQHFTRIRFDYTSELL 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 AGEPLNI---RHISITTYAR-----LKLEYIADCDK---VLTYDTPVLRDG-LKPLMDT 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 73 ESLPTAMMKRRISKATYAREDIQVLLTADLLTPYSKRRVLYLDADVYVKADLSELLDL 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 118 DLGGMVW-GAC-----IDLPEREGE-----YKKGIGMADG-----EYFNNGYL 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 DLGEEKILIGAVEDLVNSKESEFFEDFWEORAKSRILIEQGLYETPDGLKVIQSYFNNGYL 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 157 LITNKMRHRHIDFKKSCFWEYOYK---DVMQYODDILNGLEFKG-----GYCYAN 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 193 LITNKMKREBNIFKALDELLEKKKEEDYVFFPDQVNLITLFGKGFKDEIDSIARVLELP 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 204 SRENFMPTNYAFMANGFASRHTDPLLYDRTN-TAMPAVASHYCGS-AKPMHR 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 253 PRYNNQYNNVAYMYEFLFKMRKEKTLAKKLNLPTGNSISIVYGCAGTRPMWR 304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7

[illegible]

XX	PD		02-MAY-2002.	
XX	PF		29-OCT-2001; 2001WO-GB04789.	
XX	PR		27-OCT-2000; 2000GB-0026333.	
XX	PR		24-NOV-2000; 2000GB-0028727.	
XX	PR		07-MAR-2001; 2001GB-0005640.	
XX	PA	(CHIR-) CHIRON SPA.		
XX	PA	(GENO-) INST GENOMIC RES.		
XX	PI	Telford J, Masignani V, Margarit Ros XI, Grandi G, Fraser C;		
XX	PI	Tettelin H;		
XX	DR	N-PSDB; ABN69179.		
XX	PT	New Streptococcus protein for the treatment or prevention of infection		
XX	PT	or disease caused by Streptococcus bacteria, such as meningitis, and		
XX	PT	for detecting a compound that binds to the protein -		
XX	PX	Claim 1; Page 3790-3791; 4525bp; English.		
XX	CC	The invention relates to a protein (ABP25413-ABP30895) from group B		
XX	CC	Streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS		
XX	CC	(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in		
XX	CC	the specification. The proteins have antibacterial and antiinflammatory		
XX	CC	activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and		
XX	CC	antibodies that bind (I) are used in the manufacture of medicaments for		
XX	CC	the treatment or prevention of infection or disease caused by		
XX	CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.		
XX	CC	Nucleic acids encoding (I) are used to detect Streptococcus in a		
XX	CC	biological sample. (I) is used to determine whether a compound binds to		
XX	CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be		
XX	CC	used as a vaccine or diagnostic composition. The disease caused by		
XX	CC	Streptococcus that is prevented or treated may be meningitis. Nucleic		
XX	CC	acid encoding (I) may be used to recombinantly produce (I) and may be		
XX	CC	used in gene therapy. Antibodies to (I) are used for affinity		
XX	CC	chromatography, immunoassays, and distinguishing/identifying		
XX	CC	Streptococcus proteins.		
SQ		Sequence 401 AA:		
		Query Match 12.3%; Score 203.5; DB 23; Length 401;		
		Best Local Similarity 27.4%; Pred. NO. 7.7e-13;		
		Matches 74; Conservative 45; Mismatches 110; Indels 41; Gaps 11;		
OY		3 IYFAADDNVTAALCAAKSVLEAAHNDTELRFRIVLDAGISEENRAVAANLROGGNIRFTD 62		
Dd		: :		
		5 VALAADSNTLDAKALVTIKSI--CYVNRRNTFFLFPNODFEVEWRNRNRKLEPLGS-KLIN 61		
OY		63 VNPEFAGEPLNIRH---ISITTYARLKIGEYADCDCVLVDLDVDVLRDGLKLPMDDT 118		
Dd		: : : : : :		
		62 VKIYN-----DIAHLTFLTVSTWRFRLFLAYIPSS-VSLVLDSDLIYTNNFDLYIELD 115		
OY		119 LCGNMVGACIDLFVERQESYKOKIGMADGEYFPNAGVLLINLKKMRHRDIFKMSCWEHQ 178		
Dd		: :		
		116 FKGYYLAIAFKDKHKNEEG-----FNAGMLIANELMWREDELFTTLKTAE 162		
OY		179 YKDVOYOODOILNLGFKGVGYANSRENFMPTNAFAMNGFASR--HTDPYLIDRTNTA 236		
Dd		:		
		163 LHRVVKTDQSILN-----IVCH--NRWLSTLKTNNFOTDYDVYSKRINHNSYLTLENMT 215		
OY		237 MPVAVSHYCGRSAKPWHRDCTVWGAEHFTEL 266		
Dd		: : : :		
		216 PNII--HEFLTSDKPWNN-----SVARFREL 239		
RESULT 11				
ID		AAG15767		
XX		AAG15767 standard; Protein; 332 AA.		

AC AAG15767;
XX
XX 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 16147.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EPI03405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128845.
PR 16-APR-1999; 99US-0130077.
PR 19-APR-1999; 99US-0130449.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

```
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 11.9%; Score 197.5; DB 21; Length 332;
Best Local Similarity 26.3%; Pred. No. 2,6e-12;
Matches 76; Conservative 38; Mismatches 104; Indels 71; Gaps 14;

```
OY 3 IVEAADDNVAATYLCVAKGV--EAAHPDEIRPHVLDAGISEENR--AAVAANLRGGNIN 58
DB 54 VAITLDEVLRGSIASVNSILHSCVPE--VFHFHFL--AVSEETNILESIVRSVFGLKF 110
OY 59 RFLDVNPEDEPAGF-----PLNIRHISITTYARLKLGEYIADC-DKVLILDTDVL 106
DB 111 NIYDFAPETVRGSISSVQAOLEQPLN-----YARSLADLLEPCVNRVITYLSDLV 162
OY 107 VRDGLPLMDFTDIDGMWVGA---CIDLEVRQEGYKOKIGMADGEY-----YFNAGVLL 157
DB 163 VYDDIARKLWKTSLGSRILIAPEYCYANFTKYFTGFGFMSSEERFSGFRGRKPCYFNTGVWV 222
OY 158 INKKRRRHIDIFMSCEWGEQYKDWQYODODLNL-----FKGGVCANSRFNMPT 211
DB 223 IDLKKWRGGYTRIKEMMETIQRRRIYE---LGSLLPPLLVFAGHVAISHRMN----- 274
```

```
OY 212 NYAFMANGF-----ASRHTDPLYLDRNTAMPVAVSHYCGSARPKMR 253
DB 275 -----QHGLGGDNVRGSGCRDLHP-----GPVSLLLHWSGSGKFWLR 309

RESULT 12
AAG15766
ID AAG15766 standard; Protein: 346 AA.
XX
XX AAG15766;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 16146.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0122180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 23-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
```

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145195.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151308.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152263.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.9%; Score 197.5; DB 21; Length 346;
Best Local Similarity 26.3%; Pred. No. 2,7e-12;
Matches 76; Conservative 38; Mismatches 104; Indels 71; Gaps 14;
QY 3 IVEFADDNYAAYLCVAKSV--EAHPDEIRFHVLDAGISEENR--AAVAANLRGGNI 58
DB 68 VATITDVEYLRGSIANVNSLQHSVCPES-VFFHFI--AVSEETNLESLVRSVFPGLKF 124
QY 59 RFTDVNDEDFAGF-----PLNIRHISITTTARLKLGEYIADC-DKVLXLDTDLV 106
DB 125 NIYDFAPETVRGLISSVSROALEOPLN-----YARSYLADLLRCPVNRVIYLDSDLV 176

Db 7 VLADGYSYIRQIETTLKSLCYVHENDSI--FIFNÖDIPQEWFLAMKDRVGQIGNQIDVY 64

QY 58 IRFDVNPEDFAGFPLNIRHISTTYARLKGLEYIADCDKVLVLDTDVLRDGHKPLMDT 117

CC acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying

XX Sequence 398 AA;
SQ Query Match 11.9%; Score 197; DB 23; Length 398;
Best Local Similarity 28.1%; Pred. No. 3.8e-12;
Matches 73; Conservative 42; Mismatches 103; Indels 42; Gaps 11;

QY 3 IVFADADNYAAVLCAVAKSVENAHPTREIRFHVLDAGISEENRAVAANLKGGN----- 57
DB VVLAGDYSYINGIETTLKSLCVYHENLSI--FIFMODIPQEMFLAMKRVQGTGQIDV 64
QY 58 IRFDVNPEDFAGPPLNIRHISITTYARLKGEYIADCDKVLVLDPTDVLVRDGLKPLMDT 117
DB 65 KLFHDHLSPKWENKTLN--HINMYTARYFTIPOYIS-ADFTVLYLDDSDLVVTTNLDNLFQI 121

QY 118 DLGGMWVACIDLFPEREGYKQKIGMADGEYFNAGVLLINLKKRRHDIFFKSCWEVE 177
DB 122 SLDNAYLAAVPALE-----GLG--YFNAGVAVINNOQRQEN--MTIKLIE 164

QY 178 QYKDVMOYQ---DQDILNGLFKGGVCYANSRPNEMPTNYAFMANGF---ASRHTDPLYLID 231
DB 165 KNQKRIEVAANGSDQTLINRMFENQYIYLDYRYNFO-----IGFDMGAALDGHKFTFD 216

QY 232 RTNFMAPVAVSHYCGSAKFW 251
DB 217 IPITPLP-KIHHYISGIKFW 235

RESULT 15
ABP27411
ID ABP27411 standard; Protein; 312 AA.
XX AC ABP27411;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 3998.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
OS Streptococcus agalactiae.
XX MO200234771-A2.
XX PN 02-MAY-2002.
XX PD 29-OCT-2001; 2001WO-GB04789.
XX PF 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX WPI: 2002-352536/38.
XX DR N-PSDB; ABN68042.
XX PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX Claim 1; Page 3555; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.

XX Sequence 312 AA;
SQ Query Match 11.8%; Score 196; DB 23; Length 312;
Best Local Similarity 33.0%; Pred. No. 3.4e-12;
Matches 61; Conservative 22; Mismatches 70; Indels 32; Gaps 8;

QY 83 YARLKGEYIADCDKVLVLDPTDVLVRDGLKPLMDIDLGNNVAGACIDLFPERQEGYKQKI 142
DB 3 YARTIIPQLI-DAEKVLYLDDITDVLVDNLDKLFETELGDIPIAATLD----- 48

QY 143 GMADEGEYFNAGVLLINLKKRRHDIFFKSCWEVBOYKDVMOYODILNGLFKGGCYA 202
DB 49 --GDC-IYFNSGVMLINSLWMRYRVEKLEITERELDNGIFGQGVNLFLFNNMLKL 105

QY 203 NSRFNEMPTN--YAFMANGFASRHTDPLYLDRNTNAMPVAVSHYCGSAKPMHRDCTWGA 260
DB 106 EDKYNAQVGNLGAPEYENWQ-----YFDR-NPESPTII-HYCTHDKPMN---TFSS 152

QY 261 ERFTE 265
DB 153 SRFRE 157

Search completed: December 2, 2002, 11:59:01
Job time : 33.7609 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 2, 2002, 11:55:41 ; Search time 11.3901 Seconds
(without alignments)
790.458 Million cell updates/sec

Title: US-10-007-267-4

Perfect score: 1657
Sequence: 1 MDIVFAADNDYAAVLCVAAK.....KCMIDQRMKKLSARFLRKIV 306

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCRTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1657	100.0	306	1 US-08-312-387B-4	Sequence 4, Appli
2	1657	100.0	306	1 US-08-683-426-4	Sequence 4, Appli
3	1657	100.0	306	1 US-08-683-458-4	Sequence 4, Appli
4	1657	100.0	306	2 US-08-878-360-4	Sequence 4, Appli
5	1657	100.0	306	3 US-08-478-140B-4	Sequence 4, Appli
6	1657	100.0	306	4 US-09-333-412-4	Sequence 4, Appli
7	1657	100.0	306	4 US-09-338-943-4	Sequence 4, Appli
8	583	33.2	106	4 US-09-376-330-28	Sequence 28, Appli
9	280	16.9	106	4 US-09-376-330-27	Sequence 27, Appli
10	176.5	10.7	674	4 US-08-961-083-200	Sequence 200, App
11	140.5	8.5	105	4 US-09-376-330-29	Sequence 29, Appli
12	132	8.0	328	4 US-09-073-297-6	Sequence 6, Appli
13	130	7.8	106	4 US-09-376-330-25	Sequence 25, Appli
14	121.5	7.3	121	4 US-09-376-330-30	Sequence 30, Appli
15	115	6.9	110	4 US-09-376-330-24	Sequence 24, Appli
16	110.5	6.7	107	4 US-09-376-330-26	Sequence 26, Appli
17	110.5	6.7	1548	4 US-09-376-330-15	Sequence 15, Appli
18	109	6.6	110	4 US-09-376-330-23	Sequence 23, Appli
19	98	5.9	346	4 US-09-073-297-2	Sequence 2, Appli
20	91.5	5.5	115	4 US-09-376-330-20	Sequence 20, Appli
21	89.5	5.4	1327	4 US-09-376-330-2	Sequence 2, Appli
22	87	5.3	394	4 US-09-055-765-2	Sequence 2, Appli
23	87	5.3	576	1 US-08-221-817-22	Sequence 22, Appli
24	87	5.3	576	1 US-08-454-439-22	Sequence 22, Appli
25	87	5.3	576	5 PCR-US94-10487-22	Sequence 22, Appli
26	86	5.2	1447	4 US-09-376-330-17	Sequence 17, Appli
27	81	4.9	389	4 US-09-055-765-15	Sequence 15, Appli

28	80.5	4.9	437	4 US-08-961-083-70	Sequence 70, Appli
29	78.5	4.7	1019	4 US-09-434-066-23	Sequence 23, Appli
30	78	4.7	566	4 US-09-073-297-5	Sequence 5, Appli
31	78	4.7	1041	1 US-08-220-151-4	Sequence 4, Appli
32	78	4.7	1041	1 US-08-413-118-4	Sequence 4, Appli
33	78	4.7	1041	3 US-08-473-446-4	Sequence 4, Appli
34	78	4.7	1702	4 US-08-296-791-5	Sequence 5, Appli
35	78	4.7	1702	5 PCR-US95-10661A-5	Sequence 5, Appli
36	78	4.7	2802	4 US-09-542-331-1	Sequence 1, Appli
37	78	4.7	2802	4 US-09-510-791-1	Sequence 1, Appli
38	77	4.6	1493	4 US-09-376-330-16	Sequence 16, Appli
39	77	4.6	1581	4 US-08-726-320-3	Sequence 3, Appli
40	77	4.6	1581	4 US-09-208-716-3	Sequence 3, Appli
41	77	4.6	1835	3 US-08-836-325-15	Sequence 15, Appli
42	76.5	4.6	117	4 US-09-376-330-34	Sequence 34, Appli
43	76.5	4.6	1066	2 US-08-633-770A-1	Sequence 1, Appli
44	76	4.6	754	4 US-09-521-780-2	Sequence 2, Appli
45	76	4.6	754	4 US-09-521-780-3	Sequence 3, Appli

ALIGNMENTS

RESULF 1
US-08-312-387B-4
; Sequence 4, Application US/08312387B
; Patent No. 5545553

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/312,387B
FILING DATE: July 7, 1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 306 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-312-387B-4

Query Match 100.0%; Score 1657; DB 1; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.9e-171;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVFAADNDYAAVLCVAAKSYEAHPDTEIRPHVLDAGISEENRAVAANRGGINRF 60
DB 1 MDIVFAADNDYAAVLCVAAKSYEAHPDTEIRPHVLDAGISEENRAVAANRGGINRF 60
QY 61 IDVNPEDFAAGPLNIRHISITTYARLKGELYADCDKVLVLDTVLVDRGLKPLMDTDLG 120

Db 61 IDVPEDFAGFPLINRHSITTYARLKLGEYIADCCKVLYLDTVDVLRDGLKPLMDTDLG 120
121 GNNVACIDLFVERQEGYKQKIGMADGEYFNAGVLLINLKKRRHDFKMSCEWEQYK 180
121 GNNVACIDLFVERQEGYKQKIGMADGEYFNAGVLLINLKKRRHDFKMSCEWEQYK 180
181 DVMOYDODILNLFKGGVCYANSRNFEMPTNFAFANGFASRHTDPLDRTNTAMPVA 240
181 DVMOYDODILNLFKGGVCYANSRNFEMPTNFAFANGFASRHTDPLDRTNTAMPVA 240
241 VSHYCSAKPMHRCCTVWGAERFTELAGSLTTPVEEMRGLAVPTKMLQRRKKLSAR 300
241 VSHYCSAKPMHRCCTVWGAERFTELAGSLTTPVEEMRGLAVPTKMLQRRKKLSAR 300
301 FLRKTY 306
301 FLRKTY 306

RESULT 2

US-08-683-426-4
Sequence 4, Application US/08683426
Patent No. 5705367
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,426
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-0958
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 306 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-426-4

Query Match 100.0%; Score 1657; DB 1; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.9e-171;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVFAADNNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAAVANLGGGNIRF 60
1 MDIVFAADNNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAAVANLGGGNIRF 60

QY 61 IDVPEDFAGFPLINRHSITTYARLKLGEYIADCCKVLYLDTVDVLRDGLKPLMDTDLG 120
121 GNNVACIDLFVERQEGYKQKIGMADGEYFNAGVLLINLKKRRHDFKMSCEWEQYK 180
121 GNNVACIDLFVERQEGYKQKIGMADGEYFNAGVLLINLKKRRHDFKMSCEWEQYK 180
181 DVMOYDODILNLFKGGVCYANSRNFEMPTNFAFANGFASRHTDPLDRTNTAMPVA 240
181 DVMOYDODILNLFKGGVCYANSRNFEMPTNFAFANGFASRHTDPLDRTNTAMPVA 240
241 VSHYCSAKPMHRCCTVWGAERFTELAGSLTTPVEEMRGLAVPTKMLQRRKKLSAR 300
241 VSHYCSAKPMHRCCTVWGAERFTELAGSLTTPVEEMRGLAVPTKMLQRRKKLSAR 300
301 FLRKTY 306
301 FLRKTY 306

RESULT 3

US-08-683-458-4
Sequence 4, Application US/08683458
Patent No. 5798233
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,458
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-0958
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 306 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-458-4

Query Match 100.0%; Score 1657; DB 1; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.9e-171;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVFAADNNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAAVANLGGGNIRF 60
1 MDIVFAADNNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAAVANLGGGNIRF 60

QY 61 IDVNEDEAGFPLNIRHISITTYARLKGEYIADCDKYLITDVLVBDGLKPLMDTDLG 120
| | | | |
Db 61 IDVNEDEAGFPLNIRHISITTYARLKGEYIADCDKYLITDVLVBDGLKPLMDTDLG 120
QY 121 GNMVACIDLFYERDEGKOKIGMADGEYFNAGVLLINLKKRRHDIFFKMSCEWVEQYK 180
| | | | |
Db 121 GNMVACIDLFYERDEGKOKIGMADGEYFNAGVLLINLKKRRHDIFFKMSCEWVEQYK 180
QY 181 DVMQYODDILNGLFKGGVCYANSRNFNPTNYAFMANGFASRHTDPLYLDRNTAMPYA 240
| | | | |
Db 181 DVMQYODDILNGLFKGGVCYANSRNFNPTNYAFMANGFASRHTDPLYLDRNTAMPYA 240
QY 241 VSHYGSAPMHRDCTVWGAERFTELAGSLTTPPEMRGKLAVPTKCMLOMRKKLSAR 300
| | | | |
Db 241 VSHYGSAPMHRDCTVWGAERFTELAGSLTTPPEMRGKLAVPTKCMLOMRKKLSAR 300
QY 301 FLRKITY 306
| | | | |
Db 301 FLRKITY 306

RESULT 4

US-08-878-360-4
; Sequence 4, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,360
; FILING DATE: 18-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,426
; FILING DATE:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-878-360-4

Query Match 100.0%; Score 1657; DB 2; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.9e-171;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVFAADNNYAALCYVAAKSVAAHPDTEIRFHVLDAGISEENRAVAANIRGGGINRF 60
| | | | |
Db 1 MDIVFAADNNYAALCYVAAKSVAAHPDTEIRFHVLDAGISEENRAVAANIRGGGINRF 60
QY 61 IDVNEDEAGFPLNIRHISITTYARLKGEYIADCDKYLITDVLVBDGLKPLMDTDLG 120
| | | | |
Db 61 IDVNEDEAGFPLNIRHISITTYARLKGEYIADCDKYLITDVLVBDGLKPLMDTDLG 120
QY 121 GNMVACIDLFYERDEGKOKIGMADGEYFNAGVLLINLKKRRHDIFFKMSCEWVEQYK 180
| | | | |
Db 121 GNMVACIDLFYERDEGKOKIGMADGEYFNAGVLLINLKKRRHDIFFKMSCEWVEQYK 180
QY 181 DVMQYODDILNGLFKGGVCYANSRNFNPTNYAFMANGFASRHTDPLYLDRNTAMPYA 240
| | | | |
Db 181 DVMQYODDILNGLFKGGVCYANSRNFNPTNYAFMANGFASRHTDPLYLDRNTAMPYA 240
QY 241 VSHYGSAPMHRDCTVWGAERFTELAGSLTTPPEMRGKLAVPTKCMLOMRKKLSAR 300
| | | | |
Db 241 VSHYGSAPMHRDCTVWGAERFTELAGSLTTPPEMRGKLAVPTKCMLOMRKKLSAR 300
QY 301 FLRKITY 306
| | | | |
Db 301 FLRKITY 306

RESULT 5

US-08-478-140B-4
; Sequence 4, Application US/08478140B
; Patent No. 6127153
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: ROTH, STEPHEN
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,140B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-140B-4

Query Match 100.0%; Score 1657; DB 3; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.9e-171;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-09-338-943-4

Query Match 100.0%; Score 1657; DB 4; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.9e-171;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVEAADNNAAYLCVAKSVEAAHPTIEIRFHVLDAGISEENRAAVALRGGINIF 60
DB 1 MDIVEAADNNAAYLCVAKSVEAAHPTIEIRFHVLDAGISEENRAAVALRGGINIF 60
QY 61 IDVNEDEAGFPLNRHSITTYARLKGEYIADCDKVLVLDTVLVKRLPLMDTDIG 120
DB 61 IDVNEDEAGFPLNRHSITTYARLKGEYIADCDKVLVLDTVLVKRLPLMDTDIG 120
QY 121 GNMVACIDLFVEROEGYKOKIGMADGEYFNAGVLLINLKRRHDIKMSCEWEOYK 180
DB 121 GNMVACIDLFVEROEGYKOKIGMADGEYFNAGVLLINLKRRHDIKMSCEWEOYK 180
QY 181 DVMQYODDILNGLFKGVCYANSRPNMPTNYAFMANGFASRHTDPLYLDRTNTAMPYA 240
DB 181 DVMQYODDILNGLFKGVCYANSRPNMPTNYAFMANGFASRHTDPLYLDRTNTAMPYA 240
QY 241 VSHYGSAPMHRDCTVMGAERETELAGSLTVPPEMRGKLAVPPTKMLQWRKKLSAR 300
DB 241 VSHYGSAPMHRDCTVMGAERETELAGSLTVPPEMRGKLAVPPTKMLQWRKKLSAR 300
QY 301 FLRKTY 306
DB 301 FLRKTY 306

RESULT 8

US-09-376-330-28
; Sequence 28, Application US/09376330
; Patent No. 6399321

; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignard, Daniel
; APPLICANT: Bergeron, John J.M.
; APPLICANT: Thomas, David Y.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
; TITLE OF INVENTION: and nucleic acid encoding for UGCT
; FILE REFERENCE: 2139-9"US"
; CURRENT APPLICATION NUMBER: US/09/376,330
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GLYT-N, gon
US-09-376-330-28

Query Match 35.2%; Score 583; DB 4; Length 106;
Best Local Similarity 100.0%; Pred. No. 2.1e-55;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 IADCKVLYLDTVLVRDGLKPLMDTDIGNMVACIDLFVEROEGYKOKIGMADGEYF 151
DB 1 IADCKVLYLDTVLVRDGLKPLMDTDIGNMVACIDLFVEROEGYKOKIGMADGEYF 60
QY 152 NAGVLLINLKRRHDIKMSCEWEOYKDVQYODDILNGLFKG 197
DB 61 NAGVLLINLKRRHDIKMSCEWEOYKDVQYODDILNGLFKG 106

RESULT 9

US-09-376-330-27
; Sequence 27, Application US/09376330
; Patent No. 6399321

; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignard, Daniel
; APPLICANT: Bergeron, John J.M.
; APPLICANT: Thomas, David Y.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
; TITLE OF INVENTION: and nucleic acid encoding for UGCT
; FILE REFERENCE: 2139-9"US"
; CURRENT APPLICATION NUMBER: US/09/376,330
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GLUT-H, Inf
US-09-376-330-27

Query Match 16.9%; Score 280; DB 4; Length 106;
Best Local Similarity 46.7%; Pred. No. 1.1e-22;
Matches 50; Conservative 24; Mismatches 31; Indels 2; Gaps 2;

QY 92 IADCKVLYLDTVLVRDGLKPLMDTDIGNMVACIDLFV-ROEGYKOKIGMADGEY 150
DB 1 IKNIKAIYIDVDTLNSLOGLMINDITNYLACROFDIVKNAKYKTTGL-EGY 59
QY 151 FNAGVLLINLKRRHDIKMSCEWEOYKDVQYODDILNGLFKG 197
DB 60 FNAGVLLINLKRRHDIKMSCEWEOYKDVQYODDILNGLFKG 106

RESULT 10

US-08-961-083-200
; Sequence 200, Application US/08961083
; Patent No. 6159469

; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 200:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 674 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

;          TOPOLOGY:  linear
;          MOLECULE TYPE:  protein
US-08-961-083-200

```

Query Match	10.7%	Score 176.5	DB 4	Length 674
Best Local Similarity	25.1%	Pred. No. 2.6e-10		
Matches 72; Conservative	43	Mismatches 107	Indels 65	Gaps 11

QY 3 IYRADDNTAAYLCAVAAKSVEAAHPDTEIRFHVLDAGISENNRAAVANLRGGNIREID 62
 ||||: ||: ||||: | ||: ||
 Db 283 IYLAANYGYVDQVLEVIKSI--CYHNRSIRFYLIHSDFPNE-----WIKQLN 322

QY	63	VNPEDFAGEPLNIR-----HISITTYARLKGLEYIADCCKVLYLDTPLVVRDGL	11
		:	
Db	328	KRLEKFDSEINCRAVTSQISCTKSDISTYVFLRFIADPQOE-DKALYLDCCILVVTKNL	38

DQY 112 KPLMDTDLGSNMWGACIDLFVEROEGYKOKIGMA-DGEYYNAGVLLINLKKRRHDI^FK 17
| : || | : | | | | | | | : :: : :
Db 387 DDLFATDLDYPLAAVRDFG-----GRAYGOEIFNAGVLVNNAFNKENMIO 43

QY 171 MSCBWEQYKDVMOYQDDILNGLFKGVCYANSPFNEMPTNYAFMANGFASRHTDPLYL 23
 : :: | : || ||| || : : : : |
Pb 436 KLIDVTENEMDKVDADOSILNMFFHKWLELDENYNHIVTHKOFADYOI.PEGODYP--- 49

```

QY      231 DRTNAMPVAVSHYCGSAPW-----HRDCTVW-----GAERFTEL 266
          | : | | | | | | | | | | | | | | | | | | | | | |
Db      493 -----AIIHYLSHRKPKWDLAAQTYRE--VWVIYHGLE-WTEL 527

```

RESULT 11
US-09-376-330-29
; Sequence 29, Application US/09376330

GENERAL INFORMATION: *
 APPLICANT: Tessler, Daniel C.
 APPLICANT: Dignard, Daniel
 APPLICANT: Bergeron, John J. M.
 APPLICANT: Thomas, David Y.
 TITLE OF INVENTION: Method for screening for
 TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
 TITLE OF INVENTION: and nucleic acid encoding for UGGT

```
;; SOFTWARE: FastSEQ for Windows Version 3.0
;; SEQ ID NO 29
;; LENGTH: 105
```

; FEATURE:
; OTHER INFORMATION: GSPA-B. sub
US-09-376-330-29

Query Match	8.5%	Score 140.5	DB 4	Length 105
Best Local Similarity	28.0%	Pred. No. 1.3e-07		
Matches 28; Conservative	27	Mismatches 40	Indels 5	Gaps 2

Qy 97 KVLVLDITVLVRDGLKPLMBDIDLGNNVACIDLFEVERQEGYKKIKGMADGEYIFNAGVL 156
:::| | | : | | | : | | | :
Db 6 RMIIYIDCALVLEDISKLMDDIAPIYVAAVED--AGQHERLKEMNVTDTGKYFNSGIM 62

Qy 157 LINLKKRRRHDIFFKMSCEWVEY - KVMQYQDODILNGL 1944
::: ||::: : : : : || || :
Db 63 IIDFESRRKQNTTEKAVINFINHPDEFLVLHDQDALNAI 1022

RESULT 12
US-09-073-297-6
; Sequence 6, Application US/09073297

Patent No. 6255114
GENERAL INFORMATION:
APPLICANT: LIGHTNER, JONATHAN EDWARD

3

? TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES
 ?
 ? NUMBER OF SEQUENCES: 26
 ?
 ? CORRESPONDENCE ADDRESS:
 ?
 ? ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
 ? STREET: 1007 MARKET STREET
 ? CITY: WILMINGTON
 ? STATE: DELAWARE
 ? COUNTRY: U.S.A.

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE, 3.50 INCH
:
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: MICROSOFT WINDOWS 95
: SOFTWARE: MICROSOFT WORD VERSION 7.0A
: CURRENT APPLICATION DATA:

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/852615

REGISTRATION NUMBER: 41,173
REFERENCE/DOCKET NUMBER: BB-1083-B

```

; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20

```

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
ms-09-073-207-6

```

Query Match	8.0%;	Score 132;	DB 4;	Length 328;
Best Local Similarity	23.1%;	Pred. No. 5.6e-06;		
Matches	56;	Conservative 42;	Mismatches 64;	Indels 80;
				Gaps 15

95 VINSKLRIMEFV-EYERMVLDADIQVEDNIDHLEFDLDPGAFY--AVKDFCECKETSHT 151

QY 182 VMQ-----YQDQDILNGLEKGVVCYANSRNFNMPNTYAFMANGFASRHTDPLYLDR

QY 233 TTNAMPVAVSHYCCG - AKPMHRDCTVWGAERFTELAGSLTTPPEEMRGKLAVPPTKMLQ 291

Db	283	RW	284
----	-----	----	-----

RESULT 13
US-09-376-330-25
; Sequence 25, Application US/09376330

; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignard, Daniel
; APPLICANT: Bergeron, John J.M.

;	TITLE OF INVENTION: Method for screening for
;	UDP-glucose:glycoprotein glucosyltransferase (UGT)
;	TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGT) activity

3


```

; TITLE OF INVENTION: and nucleic acid encoding for UGCT
; FILE REFERENCE: 2139-9"US"
; CURRENT APPLICATION NUMBER: US/09/376,330
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GLUT-E. coli
US-09-376-330-25

```

```

Query Match          7.8%; Score 130; DB 4; Length 106;
Best Local Similarity 33.3%; Pred. No. 1.8e-06;
Matches 35; Conservative 20; Mismatches 44; Indels 6; Gaps 3;

```

```

QY 96 DKVLYIDTVLVRDGLKPLMDTDLGNNVVGACIDLFVEREGYKOKIGMADGE---YYPN 152
DB 5 DRLVYIDADVCKGDISQLHLGLNGA-VAAVVKDVEPMQE--KAVSRLSDELLGQYFN 61
QY 153 AGVLLINKKRRRHDFKMSCEWEQYKDVMOYQODDILNGLFKG 197
DB 62 SCVVYVLDLKKMADAKLTERKALSIKMSKDVNKKYKPDQDVANNVLLKG 106

```

RESULT 14

```

US-09-376-330-30
; Sequence 30, Application US/09376330
; Patent No. 6399321

```

```

; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignard, Daniel
; APPLICANT: Bergeron, John J.M.
; APPLICANT: Thomas, David Y.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGCT) activity
; FILE REFERENCE: 2139-9"US"
; CURRENT APPLICATION NUMBER: US/09/376,330
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Q48480-K. P
US-09-376-330-30

```

```

Query Match          7.3%; Score 121.5; DB 4; Length 121;
Best Local Similarity 28.3%; Pred. No. 1.8e-05;
Matches 36; Conservative 20; Mismatches 32; Indels 39; Gaps 5;

```

```

QY 96 DKVLYIDTVLVRDGLKPLMDTDLGNNVVGACIDLFVEREGYKOKIGMA-----DG 147
DB 5 DKVFLDSTVYKADLGELLDVPLGNLVAAYKDIW---EGVVKFSASASDQGMPIG 61
QY 148 EX-----YFNAGVLLINKKRRRHDF-----KMSCEWEQYKDVMOYQOD 188
DB 62 EYLQKTLNNNNPDEYFQAGIIVFNKQWVEENTFALIMRVLLAKKKYV-----FLDQ 112
QY 189 DILNGLF 195
DB 113 DIMNKVF 119

```

```

RESULT 15
US-09-376-330-24
; Sequence 24, Application US/09376330
; Patent No. 6399321

```

```

; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignard, Daniel
; APPLICANT: Bergeron, John J.M.
; APPLICANT: Thomas, David Y.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGCT) activity
; FILE REFERENCE: 2139-9"US"
; CURRENT APPLICATION NUMBER: US/09/376,330
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GALT-S. typ
US-09-376-330-24

```

```

Query Match          6.9%; Score 115; DB 4; Length 110;
Best Local Similarity 28.8%; Pred. No. 7.7e-05;
Matches 32; Conservative 17; Mismatches 36; Indels 26; Gaps 4;

```

```

QY 97 KVLVYIDTVLVRDGLKPLMDTDLGNNVVGACIDLFVEREGYKOKIG 143
DB 6 KVLVYIDADVCKGDISQLHLGLNGA-VAAVVKDVEPMQE--KAVSRLSDELLGQYFN 56
QY 144 MADGEYFNAGVLLINKKRRRHDFKMSCEWEQYKDVMOYQODDILN 192
DB 57 LVSG--YFNAGVLLIXIPLMTAENISKRAIEMLKQBEVYQRTTHDQDVLN 105

```

```

Search completed: December 2, 2002, 12:03:14
Job time : 12.3901 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 12:00:15 ; Search time 6.13314 Seconds
(without alignments)
794.504 Million cell updates/sec

Title: US-10-007-267-4

Perfect score: 1657

Sequence: 1 MDIVFAADNYAAVLCVAAK.....KCMQWRKKLSARFLRKIV 306

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1657	100.0	306	12 US-10-007-267-4	Sequence 4, Appl1
2	176.5	10.7	674	10 US-09-765-272-200	Sequence 200, App
3	157	9.5	372	10 US-09-881-752A-320	Sequence 320, App
4	132	8.0	328	10 US-09-829-482-6	Sequence 6, Appl1
5	120.5	7.2	214	10 US-09-925-999-864	Sequence 864, App
6	118.5	7.2	371	9 US-09-992-598-171	Sequence 171, App
7	118.5	7.2	371	10 US-09-989-722-171	Sequence 171, App
8	118.5	7.2	371	10 US-09-989-723-171	Sequence 171, App
9	118.5	7.2	371	10 US-09-989-729-171	Sequence 171, App
10	118.5	7.2	371	10 US-09-989-727-171	Sequence 171, App
11	118.5	7.2	371	10 US-09-989-731-171	Sequence 171, App
12	118.5	7.2	371	10 US-09-989-732-171	Sequence 171, App
13	118.5	7.2	371	10 US-09-991-073-171	Sequence 171, App
14	118.5	7.2	371	10 US-09-990-442-171	Sequence 171, App
15	118.5	7.2	371	10 US-09-991-163-171	Sequence 171, App
16	118.5	7.2	371	10 US-09-993-604-171	Sequence 171, App
17	118.5	7.2	371	10 US-09-990-456-171	Sequence 171, App
18	118.5	7.2	371	10 US-09-989-721-171	Sequence 171, App
19	118.5	7.2	371	12 US-10-052-586-146	Sequence 146, App

20	118	7.1	281	9 US-09-895-913A-282	Sequence 282, App
21	98	5.9	346	10 US-09-829-482-2	Sequence 2, Appl1
22	87	5.3	394	12 US-10-036-507-2	Sequence 2, Appl1
23	82.5	5.0	756	10 US-09-745-763-142	Sequence 142, App
24	81	4.9	389	12 US-10-036-507-15	Sequence 15, Appl1
25	81	4.9	389	12 US-10-007-693-112	Sequence 112, App
26	80.5	4.9	437	10 US-09-765-272-70	Sequence 70, Appl1
27	78.5	4.7	593	10 US-09-815-242-12306	Sequence 12306, A
28	78.5	4.7	811	10 US-09-895-382-32	Sequence 32, Appl1
29	78.5	4.7	1019	10 US-09-434-066-23	Sequence 23, Appl1
30	78.5	4.7	2310	10 US-09-995-542-10	Sequence 10, Appl1
31	78	4.7	566	10 US-09-829-482-5	Sequence 5, Appl1
32	77.5	4.7	536	10 US-09-815-242-12372	Sequence 12372, A
33	77.5	4.7	536	10 US-09-815-242-12978	Sequence 12978, A
34	77	4.6	311	9 US-09-981-876-191	Sequence 191, App
35	76.5	4.6	495	10 US-09-815-242-5040	Sequence 5040, App
36	76.5	4.6	1066	9 US-09-815-242-11716	Sequence 11716, A
37	76.5	4.6	1066	9 US-09-423-126-3	Sequence 3, Appl1
38	76.5	4.6	1066	10 US-09-280-197-5	Sequence 5, Appl1
39	74.5	4.5	316	10 US-09-765-272-156	Sequence 156, App
40	74	4.5	316	10 US-09-954-314-10	Sequence 10, Appl1
41	74	4.5	416	9 US-09-971-536-60	Sequence 60, Appl1
42	74	4.5	1169	10 US-09-771-161A-222	Sequence 222, App
43	73.5	4.4	333	9 US-09-829-482-3	Sequence 3, Appl1
44	73.5	4.4	342	10 US-09-911-150-4	Sequence 4, Appl1
45	73.5	4.4	493	10 US-09-815-242-12052	Sequence 12052, A

ALIGNMENTS

RESULT 1
US-10-007-267-4
Sequence 4, Application US/10007267
Patent No. US20020127682A1

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF OLIGOSACCHARIDES, AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klausner & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 306 amino acids
TYPE: amino acid

Db 187 FSLYEHYNESDMQIYESNNYAGFLVNLKLMRADHLEERLLNLTHQGCQCFYEPEDL 246
QY 191 INGLFKGVCYANSRFPNMP--TNVAFMANGFASRHTDPLDPTNAMPVAVSHYCGS 247
Db 247 LT-----LACY--QKVLLPIYINTHPMPAN--OKRFLP--DKKE-----IVMLHFTYV 289
QY 248 AKPMHRCCTVWGAEFRTLAGSLITVPEWRGKLAVPPTKMLQWRKKL-----S 298
Db 290 GKPW-----VLPFTFSKEMHETLTKTPRYAES 318
QY 299 ARFLRKI 305
Db 319 VKFLKQM 325

RESULT 4
US-09-829-482-6
Sequence 6, Application US/09829482
Patent No. US2002001843A1
GENERAL INFORMATION:
APPLICANT: LIGHTNER, JONATHAN EDWARD
TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/829,482
FILING DATE: 10-APR-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/852615
FILING DATE: MAY 7, 1997
ATTORNEY/AGENT INFORMATION:
NAME: MAJARIAN, WILLIAM R.
REGISTRATION NUMBER: 41,173
REFERENCE/DOCKET NUMBER: BB-1083-B
TELEPHONE: 302-992-4926
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 328 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-829-482-6

Query Match 8.0%; Score 132; DB 10; Length 328;
Best Local Similarity 23.1%; Pred. No. 8.9e-06;
Matches 56; Conservative 42; Mismatches 64; Indels 80; Gaps 15;

QY 80 ITTYARKIGETIACDKVLYIDTVLRDGLKPLMDTDLGNNWYGACIDLFVER----- 134
Db 95 VINYSKRLRWEEV-EXERVYLDADIQVFDNDHFLDLKGAFL--AVYDCFCERTWSHT 151
QY 135 ---QEGYKO---KITMADGE-----YYNAGVLLINLKKRRHDIFFMSEWEWQYKD 181
Db 152 PPDYDGYCQQRDEYAVMRELRGPPPLTFNAGMFV-----HE-----PGIGTAKD 197
QY 182 VMO-----YODDILNGLFKGVCYANSRFPNMPNTNVAFMANGFASRHTDPLYLD 232

Db 198 LLDALVTPPTFFAODPLNMFRR-----EQYKPIPVNVLV-AMLMRHPENVDDIQ 249
QY 233 TTNAPRVAVSHYCGS-AKPMHRCCTVWGAEFRTLAGSLITVPEWRGKLAVPPTKMLQ 291
Db 250 -----VKVHYCAAGSKPW-----RFT---GKEENNRE-----DIKMLVK 282
QY 292 RW 293
Db 283 RW 284

RESULT 5
US-09-925-299-864
Sequence 864, Application US/09925299
Patent No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 864
LENGTH: 214
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (138)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-864

Query Match 7.3%; Score 120.5; DB 10; Length 214;
Best Local Similarity 24.5%; Pred. No. 7.3e-05;
Matches 47; Conservative 35; Mismatches 69; Indels 41; Gaps 10;

QY 97 KVLVLDTVLRDGLKPLMDTDLGNNWYGACID-----LFVEROEG-----YK 139
Db 16 KVIYDDDVYVGDIQELDYDTLALGHAASFDDCDLPSAQDINLVGLQNTYMYGIDYR 75
QY 140 QK-----IGMADEYTFNAGVLLINLKKRRHDIFFMSEWEWQYDVMOYQDQDILNGLF 195
Db 76 KKAIRDGLISPTCSFNGVIVANNTEWKHORITKOLEKMQ--KNV-----EENLYSSL 129
QY 196 KGCVCYANSRFPNMPNTNVAFMANGFASRH---TPPLYIDRTNTAMPVAVSHYCGSARPW 251
Db 130 GGGVATSPXYLIVF-HGKISTINPLWIRHLGWNPDARISE--HFLQEAKLHWNGRHRPW 186
QY 252 -----HRDCTWV 258
Db 187 DRPSVHND--LW 196

RESULT 6
US-09-992-598-171
Sequence 171, Application US/09992598
Patent No. US20020160384A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Bolstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730PLC20
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655

PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24

PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 7.2%; Score 118.5; DB 9; Length 371;
Best Local Similarity 20.4%; Pred. No. 0.00025;
Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

QY 82 TAAKLGKGYIADCDKVLXLDIVLVDGLKPLMDTL----- 119
DB 151 TEAREFLPLVPSAKKAIYMDVDIVQGDILALYNTALKPGHAAFSDDCSASTKVIIR 210
QY 120 --GG--NMWACIDLEFVEQEGYKOKIGMADGEYFNAGVLLINKKRRHDIKFKSCW 175
DB 211 GAGNOYNYIG---YLDYKKERIRKLSKASTCSFNPQVFANLLEWKRONTNQLKRW 265
QY 176 VEOYKDVMOYDODLNGLFKSGVCYANRFFNEMPTNFAFMANGFASRT--DPLYLD- 232
DB 266 MK-----LNVEGL-----YSRTLAGSITTPPLIIV--FYQOQSTIDPMNNVH 307
QY 233 -----TNTAMPVAVSHYCGSAKPPWHRCCTWGAERFTELGLSTTVPEEW 277
DB 308 LGSSAGKRYSPQFVKAAKLLHWNHGLKPMGR-----TASYTDVWEKW 349

RESULT 7

US-09-989-722-171
Sequence 171, Application US/09989722
Patent No. US20020072067A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertlisen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kijavito, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tunas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730PLC63
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/076910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05

PRIOR APPLICATION NUMBER:	60/088212
PRIOR FILING DATE:	1998-06-05
PRIOR APPLICATION NUMBER:	60/088217
PRIOR FILING DATE:	1998-06-05
PRIOR APPLICATION NUMBER:	60/088555
PRIOR FILING DATE:	1998-06-09
PRIOR APPLICATION NUMBER:	60/088734
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088738
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088810
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088824
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088825
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088858
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/089105
PRIOR FILING DATE:	1998-06-11
PRIOR APPLICATION NUMBER:	60/089107
PRIOR FILING DATE:	1998-06-11
PRIOR APPLICATION NUMBER:	60/089150
PRIOR FILING DATE:	1998-06-12
PRIOR APPLICATION NUMBER:	60/089440
PRIOR FILING DATE:	1998-06-16
PRIOR APPLICATION NUMBER:	60/089512
PRIOR FILING DATE:	1998-06-16
PRIOR APPLICATION NUMBER:	60/089514
PRIOR FILING DATE:	1998-06-16
PRIOR APPLICATION NUMBER:	60/089523
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089538
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089599
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089599
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089600
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089617
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/089653
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/089801
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089907
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089908
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089947
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/089948
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/089952
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/090246
PRIOR FILING DATE:	1998-06-22
PRIOR APPLICATION NUMBER:	60/090252
PRIOR FILING DATE:	1998-06-22
PRIOR APPLICATION NUMBER:	60/090254
PRIOR FILING DATE:	1998-06-22
PRIOR APPLICATION NUMBER:	60/090349
PRIOR FILING DATE:	1998-06-23
PRIOR APPLICATION NUMBER:	60/090355
PRIOR FILING DATE:	1998-06-23
PRIOR APPLICATION NUMBER:	60/090429
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090431
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090435
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090444
PRIOR FILING DATE:	1998-06-24

PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090445
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090472
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090535
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090540
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090542
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090557
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090676
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090678
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090690
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090694
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090695
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090696
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090862
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090863
PRIOR FILING DATE:	1998-06-26
PRIOR APPLICATION NUMBER:	60/091360
PRIOR FILING DATE:	1998-07-01
PRIOR APPLICATION NUMBER:	60/091478
PRIOR FILING DATE:	1998-07-02
PRIOR APPLICATION NUMBER:	60/091544
PRIOR FILING DATE:	1998-07-01
PRIOR APPLICATION NUMBER:	60/091519
PRIOR FILING DATE:	1998-07-02
PRIOR APPLICATION NUMBER:	60/091626
PRIOR FILING DATE:	1998-07-02
PRIOR APPLICATION NUMBER:	60/091633
PRIOR FILING DATE:	1998-07-02
PRIOR APPLICATION NUMBER:	60/091978
PRIOR FILING DATE:	1998-07-07
PRIOR APPLICATION NUMBER:	60/091982
PRIOR FILING DATE:	1998-07-07
PRIOR APPLICATION NUMBER:	60/092182
PRIOR FILING DATE:	1998-07-09

Query Match	7.2%;	Score 118.5;	DB 10;	Length 371;
Best Local Similarity	20.4%;	Pred. No. 0.00025;		
Matches 48;	Conservative 38;	Mismatches 74;	Indels 75;	Gaps 9;

QY	82	TVTLKLTGEVTAADCKVLITDPTLVMDGLKPLMTDTL-----	119
Db	151	TFARFYLPILPSSAKKATIVGVGDILALTYNLKGGHAAAFSEDCSDASTKVIR	210
QY	120	--GC--MVGACIDLFEPRQEGYVKQTGMADGEYFPAGVLLINLKRRRHDIFFKSCWE	175
Db	211	GAGQNYWIG----YLDYKKERRIKLSMKASTCSFHPGFVANULTPMKRQNITNOLEKW	265
QY	176	VEQKYDWQYODDILINGLEFGKGVCYANSRNFEMPTVAEAMANGFAFSRHT--DPYLDR-	232
Db	266	MK-----LNVEGCL-----YSRTLAGSITTPPLLIV--FYQDHSTIDPMNVVRH	307
QY	233	-----TNTMPAVALSHYCASAAPMHNRDCTVWGAEKEFTELAGSLTVYPDEW	277
Db	308	LGSASGRKRYSPQFVKAALKLHMNHDLKPWR-----TASYTDYWBRK	349

RESULT 8
US-09-989-723-171
; Sequence 171, Application US/09989723
; Patent No. US20020072092A1

GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Garber, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C62
CURRENT APPLICATION NUMBER: US/09/989,723
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326

PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24

PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 7.2%; Score 118.5; DB 10; Length 371;
Best Local Similarity 20.4%; Pred. No. 0.00025;
Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

82 TYARLKGIVYINDCKVLYLDIVRODLKPLMDTL----- 119
151 TFAREYLPILVSAKAIYMDDVLYOGDILALYNALPGHAAAFSEDCDSASTKRVIR 210
120 --GG--NMVACIDLFVERQESYKOKIGMADEYFYNAGVLLINLKKMRHDIFFKMSCEM 175
211 GAGNGNYNYG-----YLDYKKERIKRLSKASTCSNPGVFYANLTFEMKRONITNDEKW 265
176 VEOYKDVMOYODDILNGLFKGCVCYANSRNFNPMPTNAFMKANGASRHT--DPLYLDR- 232
266 MK-----LNVEEG-----YSRLAGSITTPPLLIY--FYQGSTIDPMNVNRH 307
233 -----TNTAMPVANSYHSGSAKPPHRCCTYWGAEFTELASGLTFVDEEM 277
308 LGSSAGKRYSPQFVAKAALLHWNHGLKPMGR-----TASYTDVWEKW 349

RESULT 9
US-09-989-279-171
Sequence 171, Application US/09989279
Patent No. US20020072496A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paonli, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730PLC56
CURRENT APPLICATION NUMBER: US/09/989,279
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04

;; PRIOR APPLICATION NUMBER: 60/088030
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088033
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088167
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088655
;; PRIOR FILING DATE: 1998-06-09
;; PRIOR APPLICATION NUMBER: 60/088734
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088738
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088742
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088810
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088824
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088826
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088858
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088861
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088876
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089105
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089440
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089512
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089514
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089532
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089538
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089599
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089600
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089907
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089908
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089947
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/089948
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/089952
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/090246
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090252
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090254
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090349

;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090355
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090429
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090431
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090435
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090444
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090472
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090535
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090540
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090542
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090676
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090678
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090690
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090694
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090695
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090696
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090862
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091478
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091544
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091626
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091633
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091978
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/092182
;; PRIOR FILING DATE: 1998-07-09

Query Match 7.2%; Score 118.5; DB 10; Length 371;

Best local Similarity 20.4%; Pred. No. 0.00025;

Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

OY 82 TVARLKGXYINDCDVVLVDVLRDGLKPLMDRL----- 119
Db 151 TRARFLPLIVSAAKAIIMDDVIVGDIILALYNALPGRHAAASEDCDSASTVYIR 210
OY 120 --GG--NWGACIDLFVERQEGYKORIGMADGEYFNAGVLLINLKKWRHDIKNSCEM 175
Db 211 GAGNOQNYIG-----YLDYKKERIRKLSKASKSCSPNGVFYANLUEWKRONTNLEKM 265
OY 176 VEDQYKVWQYODDILNGLFKGCVGYANSRFPNPNIVAFMANGFASRRT--DPLYLDR- 232
Db 266 MK-----LNVEBGL-----YSRTLAGSITPPLIV--FYQOHSITDPMNVBH 307

QY 233 -----TNTAMPVASHYCGSAKPMHROCTVWGAEERFELASLTTPPEM 277
DB 308 LGSSACKRYSPOVRKAKLHMNGHLKPNCR-----TASYDVMWKW 349
RESULT 10
US-09-989-727-171
Sequence 171, Application US/09989727
Patent No. US20020072497A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kjaavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P27301C65
CURRENT APPLICATION NUMBER: US/09/989,727
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026

/? PRIOR FILING DATE: 1998-06-04
/? PRIOR APPLICATION NUMBER: 60/088028
/? PRIOR FILING DATE: 1998-06-04
/? PRIOR APPLICATION NUMBER: 60/088029
/? PRIOR FILING DATE: 1998-06-04
/? PRIOR APPLICATION NUMBER: 60/088030
/? PRIOR FILING DATE: 1998-06-04
/? PRIOR APPLICATION NUMBER: 60/088033
/? PRIOR FILING DATE: 1998-06-04
/? PRIOR APPLICATION NUMBER: 60/088326
/? PRIOR FILING DATE: 1998-06-04
/? PRIOR APPLICATION NUMBER: 60/088167
/? PRIOR FILING DATE: 1998-06-05
/? PRIOR APPLICATION NUMBER: 60/088202
/? PRIOR FILING DATE: 1998-06-05
/? PRIOR APPLICATION NUMBER: 60/088212
/? PRIOR FILING DATE: 1998-06-05
/? PRIOR APPLICATION NUMBER: 60/088217
/? PRIOR FILING DATE: 1998-06-05
/? PRIOR APPLICATION NUMBER: 60/088655
/? PRIOR FILING DATE: 1998-06-09
/? PRIOR APPLICATION NUMBER: 60/088734
/? PRIOR FILING DATE: 1998-06-10
/? PRIOR APPLICATION NUMBER: 60/088738
/? PRIOR FILING DATE: 1998-06-10
/? PRIOR APPLICATION NUMBER: 60/088742
/? PRIOR FILING DATE: 1998-06-10
/? PRIOR APPLICATION NUMBER: 60/088810
/? PRIOR FILING DATE: 1998-06-10
/? PRIOR APPLICATION NUMBER: 60/088824
/? PRIOR FILING DATE: 1998-06-10
/? PRIOR APPLICATION NUMBER: 60/088826
/? PRIOR FILING DATE: 1998-06-10
/? PRIOR APPLICATION NUMBER: 60/088858
/? PRIOR FILING DATE: 1998-06-11
/? PRIOR APPLICATION NUMBER: 60/088861
/? PRIOR FILING DATE: 1998-06-11
/? PRIOR APPLICATION NUMBER: 60/088876
/? PRIOR FILING DATE: 1998-06-11
/? PRIOR APPLICATION NUMBER: 60/089105
/? PRIOR FILING DATE: 1998-06-12
/? PRIOR APPLICATION NUMBER: 60/089440
/? PRIOR FILING DATE: 1998-06-16
/? PRIOR APPLICATION NUMBER: 60/089512
/? PRIOR FILING DATE: 1998-06-16
/? PRIOR APPLICATION NUMBER: 60/089514
/? PRIOR FILING DATE: 1998-06-16
/? PRIOR APPLICATION NUMBER: 60/089532
/? PRIOR FILING DATE: 1998-06-17
/? PRIOR APPLICATION NUMBER: 60/089538
/? PRIOR FILING DATE: 1998-06-17
/? PRIOR APPLICATION NUMBER: 60/089598
/? PRIOR FILING DATE: 1998-06-17
/? PRIOR APPLICATION NUMBER: 60/089599
/? PRIOR FILING DATE: 1998-06-17
/? PRIOR APPLICATION NUMBER: 60/089600
/? PRIOR FILING DATE: 1998-06-17
/? PRIOR APPLICATION NUMBER: 60/089653
/? PRIOR FILING DATE: 1998-06-17
/? PRIOR APPLICATION NUMBER: 60/089801
/? PRIOR FILING DATE: 1998-06-18
/? PRIOR APPLICATION NUMBER: 60/089907
/? PRIOR FILING DATE: 1998-06-18
/? PRIOR APPLICATION NUMBER: 60/089908
/? PRIOR FILING DATE: 1998-06-18
/? PRIOR APPLICATION NUMBER: 60/089947
/? PRIOR FILING DATE: 1998-06-19
/? PRIOR APPLICATION NUMBER: 60/089948
/? PRIOR FILING DATE: 1998-06-19
/? PRIOR APPLICATION NUMBER: 60/089952
/? PRIOR FILING DATE: 1998-06-19
/? PRIOR APPLICATION NUMBER: 60/090246
/? PRIOR FILING DATE: 1998-06-22

PRIOR APPLICATION NUMBER: 60/090252
 PRIOR FILING DATE: 1998-06-22
 PRIOR APPLICATION NUMBER: 60/090254
 PRIOR FILING DATE: 1998-06-22
 PRIOR APPLICATION NUMBER: 60/090349
 PRIOR FILING DATE: 1998-06-23
 PRIOR APPLICATION NUMBER: 60/090355
 PRIOR FILING DATE: 1998-06-23
 PRIOR APPLICATION NUMBER: 60/090429
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090431
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090435
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090444
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090445
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090472
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090535
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090540
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090542
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090557
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090676
 PRIOR FILING DATE: 1998-06-25
 PRIOR APPLICATION NUMBER: 60/090678
 PRIOR FILING DATE: 1998-06-25
 PRIOR APPLICATION NUMBER: 60/090690
 PRIOR FILING DATE: 1998-06-25
 PRIOR APPLICATION NUMBER: 60/090694
 PRIOR FILING DATE: 1998-06-25
 PRIOR APPLICATION NUMBER: 60/090695
 PRIOR FILING DATE: 1998-06-25
 PRIOR APPLICATION NUMBER: 60/090696
 PRIOR FILING DATE: 1998-06-25
 PRIOR APPLICATION NUMBER: 60/090862
 PRIOR FILING DATE: 1998-06-26
 PRIOR APPLICATION NUMBER: 60/090863
 PRIOR FILING DATE: 1998-06-26
 PRIOR APPLICATION NUMBER: 60/091360
 PRIOR FILING DATE: 1998-07-01
 PRIOR APPLICATION NUMBER: 60/091478
 PRIOR FILING DATE: 1998-07-02
 PRIOR APPLICATION NUMBER: 60/091544
 PRIOR FILING DATE: 1998-07-01
 PRIOR APPLICATION NUMBER: 60/091519
 PRIOR FILING DATE: 1998-07-02
 PRIOR APPLICATION NUMBER: 60/091626
 PRIOR FILING DATE: 1998-07-02
 PRIOR APPLICATION NUMBER: 60/091633
 PRIOR FILING DATE: 1998-07-02
 PRIOR APPLICATION NUMBER: 60/091978
 PRIOR FILING DATE: 1998-07-07
 PRIOR APPLICATION NUMBER: 60/091982
 PRIOR FILING DATE: 1998-07-07
 PRIOR APPLICATION NUMBER: 60/092182
 PRIOR FILING DATE: 1998-07-09

[illegible]

```

Db      211 GAGNGYNIIG-----YLDYKKERIRKLSMKASTCSFNGVAVLTEMKQNTINQLEKN 265

Oy      176 VEQYEDVMQYODDOLLNLFKGGCYCANSRNPMTNAPAFANGFASARH--PLPYLDR- 232
Db      266 MK-----LNNEBSL-----YSLTLGSLITPLLY--FYOOHSLIDPMVNRH 307

Oy      233 -----TNTAMPVAVSHYCSAKFMHDDCTVGAERFTELAGSLITVPEEW 277
Db      308 LGSAGKRYSDQFYKAAKLILMHNLGKMWGR-----TAYTIDWEKN 349

```

```

1 RESULT 11
2 US-09-989-731-171
3 : Sequence 171. Application us/09989731
4 : Patent No. US20020103125A1
5 :
6 : GENERAL INFORMATION:
7 :
8 : APPLICANT: Ashkenazi, Avi J.
9 : APPLICANT: Baker, Kevin P.
10 : APPLICANT: Botstein, David
11 : APPLICANT: Desnovers, Luc
12 : APPLICANT: Eaton, Dan L.
13 : APPLICANT: Ferrara, Napoleone
14 : APPLICANT: Fong, Sherman
15 : APPLICANT: Gerber, Hanspeter
16 : APPLICANT: Gerritsen, Mary E.
17 : APPLICANT: Goddard, Audrey
18 : APPLICANT: Godowski, Paul J.
19 : APPLICANT: Grimaldi, J. Christopher
20 : APPLICANT: Gurney, Austin L.
21 : APPLICANT: Kijavini, Ivar J.
22 : APPLICANT: Napier, Mary A.
23 : APPLICANT: Pan, James
24 : APPLICANT: Paoni, Nicholas F.
25 : APPLICANT: Roy, Margaret Ann
26 : APPLICANT: Stewart, Timothy A.
27 : APPLICANT: Tumas, Daniel
28 : APPLICANT: Watanabe, Colin K.
29 : APPLICANT: Williams, P. Mickey
30 : APPLICANT: Wood, William I.
31 : APPLICANT: Zhang, Zemin
32 :
33 : TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
34 : FILE OF INVENTION: Acids Encoding the Same
35 : FILE REFERENCE: P27301C70
36 :
37 : CURRENT APPLICATION NUMBER: US/09/989,731
38 :
39 : CURRENT FILING DATE: 2001-11-20
40 :
41 : PRIOR APPLICATION NUMBER: 60/049787
42 :
43 : PRIOR FILING DATE: 1997-06-16
44 :
45 : PRIOR APPLICATION NUMBER: 60/062250
46 :
47 : PRIOR FILING DATE: 1997-10-17
48 :
49 : PRIOR APPLICATION NUMBER: 60/065186
50 :
51 : PRIOR FILING DATE: 1997-11-12
52 :
53 : PRIOR APPLICATION NUMBER: 60/065311
54 :
55 : PRIOR FILING DATE: 1997-11-13
56 :
57 : PRIOR APPLICATION NUMBER: 60/066770
58 :
59 : PRIOR FILING DATE: 1997-11-24
60 :
61 : PRIOR APPLICATION NUMBER: 60/075945
62 :
63 : PRIOR FILING DATE: 1998-02-25
64 :
65 : PRIOR APPLICATION NUMBER: 60/078910
66 :
67 : PRIOR FILING DATE: 1998-03-20
68 :
69 : PRIOR APPLICATION NUMBER: 60/083322
70 :
71 : PRIOR FILING DATE: 1998-04-28
72 :
73 : PRIOR APPLICATION NUMBER: 60/084600
74 :
75 : PRIOR FILING DATE: 1998-05-07
76 :
77 : PRIOR APPLICATION NUMBER: 60/087106
78 :
79 : PRIOR FILING DATE: 1998-05-28
80 :
81 : PRIOR APPLICATION NUMBER: 60/087607
82 :
83 : PRIOR FILING DATE: 1998-06-02
84 :
85 : PRIOR APPLICATION NUMBER: 60/087609
86 :
87 : PRIOR FILING DATE: 1998-06-02
88 :
89 : PRIOR APPLICATION NUMBER: 60/087759
90 :
91 : PRIOR FILING DATE: 1998-06-02
92 :
93 : PRIOR APPLICATION NUMBER: 60/087827
94 :
95 : PRIOR FILING DATE: 1998-06-03
96 :

```

;; PRIOR APPLICATION NUMBER: 60/088021
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088025
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088026
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088028
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088029
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088030
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088033
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088167
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088655
;; PRIOR FILING DATE: 1998-06-09
;; PRIOR APPLICATION NUMBER: 60/088734
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088738
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088742
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088810
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088824
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088826
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088858
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088861
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088876
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089105
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089440
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089512
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089514
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089532
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089538
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089599
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089600
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089907
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089908
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089947
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/089948

;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/089952
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/090246
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090252
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090254
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090349
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090355
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090429
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090431
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090435
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090444
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090472
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090535
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090540
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090542
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090676
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090678
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090690
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090694
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090695
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090696
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090862
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091478
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091544
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091626
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091633
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091978
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/092182
;; PRIOR FILING DATE: 1998-07-09

Query Match 7.2%; Score 118.5; DB 10; Length 371;
Best Local Similarity 20.4%; Pred. No. 0.00025;
Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

QY 82 TYARKLGEYIADCDKVLXLDITDVILVRDGLKPLMDTDL----- 119

Db 151 TFAAFYLLVPSAKKAIYMDDDVIVCGDILALYNTALKPGHAAAFSEDCDSATKVVR 210
Qy 120 --GG--NMVAGACIDFVERQECYKOKIGMADEYFVFNAGVLLINKKRRRHIFKMSCEW 175
Db 211 GAGNGYNYIG----YLDYKKEIRIKLSMKASTCSFNGCVFVANTLEKRONITMOLKRW 265
Qy 176 VEQYVDVQYODODILNGLFKGVCYANSRFFNPTNYAFMANGFASRHT--DPLYLDR- 232
Db 266 MK-----LWNEBGL-----YSRTLAGSITTPPLIIV--FYQCHSTIDPMNNVRH 307
Qy 233 -----TNTAMPVAVSHYCGSAKPMWRDCTVWGAERFTLAGSLTTVPEEW 277
Db 308 LGSAGKRRYSPQFVKAAKLHMNGHLKPMGR-----TASYTDVWEKW 349
RESULT 12
US-09-989-732-171
Sequence 171, Application US/09989732
Patent No. US20020123463A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Bolstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zenlin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2730P1C37
CURRENT APPLICATION NUMBER: US/09/989,732
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609

PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18

Query Match	Similarity	20.4%	Pred. No. 0.00025	Best Local	48; Conservative	38; Mismatches	74; Indels	75; Gaps
7.2%	Score 118.5	DB 10;	Length 371;					
Db	151	TFARHYLPILVPSAKKAIYMDDDVIVQDIDLALYNTALKPGHAAAFSEDCDSASTKVYIR	210					
Qy	120	--GC--NMVACIDLFVERQEGYKOKICGMADGEYIFNAGVLLINKKRRHIDIFKMSCW	175					
Db	211	GAGNOYNYTG-----YLDKKERKERIKLMSKASTCSFNGFVAVANTLEWKRONITQLEKW	265					
Qy	176	VEQYKDVQYQDDQDILNLFKGVGYANSRFNFMPTNYAFMANGFASRHT--DPLYLDR	232					
Db	266	MK-----LNVEEL-----YSTRLAGSITTPILLIV--EYQHSITDIPMNVRH	307					
Qy	233	-----TNTAPVAVSHYCSAKPMHRDCTVWGAERFTELGSITVPEEW	277					
Db	308	LGSSAGKRYISQFVFAAKLHMHNGHLKRWGR-----TASTVDWEKW	349					

[illegible]

```

; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match          7.2%; Score 118.5; DB 10; Length 311;
Best Local Similarity 20.4%; Pred. No. 0.00025;
Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

QY 82 TYARLKGIVYADCKVLYDLDVLRDGLKPLMDRL----- 119
DB 151 TFAHYLPPLVPSAKAATYMDVYQGDILALVNTALPGHAAAFSECCDSASTKVIR 210
QY 120 --G--NWAGACIDLFVEREQEYKQKIGMADEYFFNAGVLLINKKRRHDIKMSCFM 175
DB 211 GAGNGYNYTG-----YLDYKKERIRLSKASTCSFNPGEFVANLTKMKRQNTNLEKM 265
QY 176 VEOYQDVMOYQDODILINGLFKGGVCYANSRFPNPTNYAFMANGASRHT--DPLYLR- 232
DB 266 MK-----LNVEEGL-----YSRLAGSITTPPLIY--FYQOHSITDPMNVNH 307
QY 233 -----TNTAMPVAVSHYCGSAKPMHROCTVWGAEFTELASLTTPPEEW 277
DB 308 LGSSAGKRYSPQVKAAKLLHMNGHLKPMGR-----TASYTDVWEKW 349

RESULT 14
US-09-990-442-171
; Sequence 171. Application US/09990442
; Patent No. US20020132252A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pad, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC8
; CURRENT APPLICATION NUMBER: US/09/990,442
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
```

```

; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
```



```

; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

```

```

Query Match 7.2%; Score 118.5; DB 10; Length 371;
Best Local Similarity 20.4%; Fred. No. 0.00025;
Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

```

```

Qy 82 TYARLKLEIYIADCKVLYLDTDLVLRDGLKPLMDTDL----- 119
Db 151 TPARFYLPILVPSAKKAIYMDDDIIVOGDILALYNTALKPGHAAAFSEDCDSASTKVIR 210
Qy 120 --GG--NWGACIDLFVERQEGYKQKIGMADGEYFNAGVLLINLKKWRHDIFKMSCEW 175
Db 211 GAGNGYNYIG-----YLDYKKERIRKLSMKASTCSFNPGEVFVANLLEWKRONITNQLERK 265
Qy 176 VBOYDVWQYQODLNLGFKGVCYANSRFEFMPNTAFMANGFASRHT--DPLYLDR- 232
Db 266 MK-----LNVEEGI-----YSRLAGSITTPPLLIY--FYQOHSITIDPMNNVRH 307
Qy 233 -----TWTAMPVAVSHYCSGSAKPMHRDCTVWGAERFTELASLTTPVEEW 277
Db 308 LGSAGKRYSPQFVAAKILHNMGNHLPWGR-----TASYTDVWEKW 349

```

```

Search completed: December 2, 2002, 12:16:01
Job time : 7.13314 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 13.5805 Seconds
(Without alignments)
2166.126 Million cell updates/sec

Title: US-10-007-267-4

Perfect score: 1657

Sequence: 1 MDIVFAADNVAAYLCVAAR.....KCMLOEWRKKLSARFLRKIIY 306

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	349	21.1	220	2 C64146	hypothetical prote
2	252	15.2	286	2 S16423	general stress pro
3	239	14.4	406	2 H95205	glycosyl transfera
4	233	14.1	398	2 G95205	glycosyl transfera
5	220	13.3	335	2 H91191	hypothetical prote
6	220	13.3	335	2 A86039	probable LPS biosy
7	210	12.7	337	2 AH0972	lipopolysaccharide
8	201.5	12.2	404	2 G95206	glycosyl transfera
9	191	11.5	339	2 C642982	lipopolysaccharide
10	190.5	11.5	346	2 T01514	lipopolysaccharide
11	190	11.5	517	2 A71707	lipopolysaccharide
12	186	11.2	361	2 C86157	hypothetical prote
13	184	11.1	572	2 T51525	hypothetical prote
14	180	10.9	338	2 S47847	lipopolysaccharide
15	179.5	10.8	814	2 A85206	glycosyl transfera
16	177	10.7	337	2 G86038	probable LPS biosy
17	177	10.7	337	2 F91191	hypothetical prote
18	176.5	10.7	696	2 D95206	hypothetical prote
19	171.5	10.4	357	2 T48055	hypothetical prote
20	170	10.3	20	2 G95922	probable glycosyl
21	168	10.1	336	2 A10972	lipopolysaccharide
22	162.5	9.8	341	2 T46153	hypothetical prote
23	159.5	9.6	394	2 A71963	probable lipopolys
24	157	9.5	372	2 G64539	lipopolysaccharide
25	157	9.5	390	2 F96723	hypothetical prote
26	156.5	9.4	658	2 T05655	hypothetical prote
27	153	9.2	589	2 F86202	hypothetical prote
28	152	9.2	393	2 T00647	glycosyl transfera
29	148.5	9.0	161	2 E97921	conserved hypother

30	144	8.7	537	2 T49162	hypothetical prote
31	139	8.4	673	2 T47905	hypothetical prote
32	136.5	8.2	582	2 F84807	hypothetical prote
33	134	8.1	335	2 G96607	probable galactino
34	133	8.0	361	2 T06009	hypothetical prote
35	132	8.0	328	2 T07610	W5176 protein - ri
36	131.5	7.9	363	2 A86267	hypothetical prote
37	130.5	7.9	377	2 E21968	probable lipopolys
38	127	7.7	373	2 F71884	probable lipopolys
39	127	7.7	921	2 F84593	hypothetical prote
40	124	7.5	61	2 A64005	hypothetical prote
41	123.5	7.5	429	2 T31001	hypothetical prote
42	122.5	7.4	344	2 A84912	probable galactino
43	117	7.1	528	2 F86226	hypothetical prote
44	115.5	7.0	528	2 D84903	hypothetical prote
45	112.5	6.8	334	2 H96629	hypothetical prote

ALIGNMENTS

RESULT 1

C64146 Hypothetical protein HI0259 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999

C:Accession: C64146

R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage

, D.M.; Brandon, R.C.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman

, Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: C64146

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-220 <TIGR>

A:Cross-references: GB:I42023; TIGR:HI0259; GB:U32711; NID:g1573214; PIDN:AA21924.1;

Query Match 21.1%, Score 349, DB 2; Length 220;

Best Local Similarity 40.2%, Pred. No. 3.2e-23;

Matches 72; Conservative 39; Mismatches 62; Indels 6; Gaps 5;

Qy	1	MDIVFAADNVAAYLCVAAR	5EAHHPDTEIRFHYLDAGISENRAAVALRGCGNIR	59
Db	39	MNIFSSDHYIAPYLAIVSIF	ITIKNP-KTINFTLDKINQENK-TIINNLAAYSCV	96
Qy	60	-FIDVNPEDFAGFPLNIRHIS	ITTYARLKGFIADCDKRVYLDTVLVVRDGLKPLMDTD	118
Db	97	FFLPYCESDFQFPRFTIDYI	SLATYARLNLTKYINIKAIYIDVTLTNSLSLOELMND	156
Qy	119	LGGMVAGACIDLFVE-ROEGYK	QKGTGMADGERYFNAQVLLNLKKRRHDIFFKSCENV	176
Db	157	ITNYLIAACRDTFIDVKN	EAYKKTIGL-EGSYFNAGILLINLNKREBNIFOKSINNM	214

RESULT 2

S16423 general stress protein gspa - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 06-Jan-1994 #sequence_revision 24-Feb-1995 #text_change 08-Dec-2000

C:Accession: S16423; S39667; P50148; A69638

R:Glaser, P.; Kunst, F.; Debarbouille, M.; Vertes, A.; Danchin, A.; Dedonder, R.

DNA Seq. 1, 251-261, 1991

A:Title: A gene encoding a tyrosine tRNA synthetase is located near sacs in Bacillus

A:Reference number: S16421; MUID:92216127; PMID:1806041

A:Accession: S16423

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 7-286 <GI2>

A:Cross-references: EMBL:X52480; NID:940236; PIDN:CAA36721.1; PID:940239

R:Glaser, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionesc

[illegible]

RESULT 5

hypothetical protein ECS4504 [imported] - *Escherichia coli* (strain O157:H7, substrain R)
C:Species: *Escherichia coli*
C:Date: 18-Jul-2001 #sequence-revision 18-Jul-2001 #text-change 03-Aug-2001
C:Accession: H91191
R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: H91191
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-335 <HAY>
A:Cross-references: GB:BA000007; PIDB:BA37927.1; PID:G1363979; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS4504
C:Superfamily: 1lipopolysaccharide glucosyltransferase I

Query Match	13.3%	Score 220;	DB 2;	Length 335;
Best Local Similarity	26.98;	Pred. No. 1.2e-11;		
Matches 73;	Conservative 45;	Mismatches 109;	Indels 44;	Gaps 13;

```

OY      1 MDIYFADDDNNAALCYAAKVEEAHHDTEIRFHYLDAGISEENK-----AAVAANDRG 55
         :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      27 LDIAGIDRNLFECGGAIAISILNNREISCEFFEPDYIDSDKDKYFSDIAQOYSNR-- 84
         :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      56 GNIREIDVNPEDFAGFPLINIHISITTYARLKLEEI-ACCDKVLVDTVLVARDGKPL 114
         :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      85 --INTYINCDKIKLSLP-STKNMYIATYAFRIIADYFIRHHEKILYLDADIDACKSGIKEL 141
         :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      115 WDTLDGGMWVGACIDLEVERO-EGYKQKIG-----MADGEYFNAGVLLINLKKRRHD 167
         :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      142 LDYOFSTEJIAAVY--AERDEVWOMONASVLTTPOLASG--YFNAGFLLINIDENMLNN 196
         :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      168 IFKMSCE-----WVEQYKIDVMOYODDILNGLEFKGGVCYANSREMPNVAFMANGA 221
         :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      197 ISSKAIEMLRBDPMYSK-----ITHDDVYLVNVLNGKVKYFISEKYN---TRYISINTE-LK 248
         :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      222 SRHTDPLYLDRTNTRAMPVAVSHYCSGAKPMH 352
         :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      249 DKVDNPVND-----TVEITHYGPKRPMH 272
         :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

RESULT 6

Probable lps biosynthesis enzyme waaI [Imported] - Escherichia coli (strain O157:H7, su
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: A86039
R:Perena, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanita, E.; Potamoustis, K.; Apodaca
Nature 409, 529-533, 2001
;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: AB6039
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-335 <STO>
A:Cross-references: GB:AE005174; NID:g12518362; PIDN:AAG58773.1; GSPDB:GN00145; UWGP
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: waaI
C:Superfamily: lipopolysaccharide glucosyltransferase I

Query Match 13.3%; Score 220; DB 2; Length 335;
Best Local Similarity 26.9%; Pred. No. 1.2e-11;
Matches 73; Conservative 45; Mismatches 109; Indels 44; Gaps 13;

Query Match	13.3%	Score 220;	DB 2;	Length 335;
Best Local Similarity	26.9%;	Pred. No.1.2e-11;		
Matches 73;	Conservative 45;	NonMatches 109;	Indels 44;	Gaps 13;
QY	1 MDIVFAADNDYAAALCYCAKSAVEAAHPDIEIFEHYLDAGISENR---- <td>55</td> <td></td> <td></td>	55		
DB	27 LDIAEGIDNPLFGCGVALASTILNNREISCFEHFTDIYSKDKLYFSDLKQYNSR--	84		
QY	56 GNIRFDVNPEDFAGFPLRIHISTYVARLKGEYI-ADCKVLYLDPDVAVRQIKPL	114		
DB	85 --INIVVICDCLKSLP-STKNWYATYTFRIIADYFYHKHEKILYLDADICKSGIKPL	141		
QY	115 WPDTLGNNVGACIDLEVERO-EGYKOKIG-----MADGEYFNAGVLLINKWRHD	167		
DB	142 LDYGFSTNIAAV--AARDYEMWNNKRSVLTTPQLASG--YFNAGFLIINDENLNLN	196		
QY	168 IFRMSE-----WEQYKDVMOQYODDILNGLFKGGVCYANSRFNEMPTNAYFANGFA	221		
DB	197 ISSKALEMRDPPDWSK-----ITHLDQDVLNVLNKGKVFISEKNV--TRSSINVE-LK	248		
QY	222 SKRTDPLYLDRNTNTAMPVAVSHYCSGSAKPMH	252		
DB	249 DKVDNVPVND-----TFVIFHYGVGTKEPMH	272		

RESULT 7

l1pPolysaccharide 3-alpha-galactosyltransferase (EC 2.4.1.44) - *Salmonella enterica*
 C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhl
 A>Note: this species has also been called *Salmonella typhi*
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AH0972
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
 , S.; Moule, S.; O'Gaara, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AH0972
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-337 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD03276.1; PID:916504897; GSPDB:GN00176
 C:Genetics:
 A:Gene: waat
 C:Superfamily: l1pPolysaccharide glucosyltransferase I
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match	12.7%;	Score 210;	DB 2;	Length 337;
Best Local Similarity	25.9%;	Pred. No. 9.5e-11;		
Matches	72;	Conservative 47;	Mismatches 111;	Indels 48;
				Gaps 13

QY 2 DIVEAADONNYAYVLCVAAKSVAAAHDPDEIREHYLD--AGISEENRAAVNAANLNGGKIR 59
DB 29 NAYGIDKNLEFGCGVSVLASVUIAAREKALAFHYTFDFGEGDOQRDALAKOYATOIV 88
QY 60 FIDVNPEDPAGEPLNIRHISTTARLKGEYIAD-CDKVLYTIDTVLVGRDGLPWDTD 118
DB 89 YL-IDCERIKSLP-STKMWYATAYFRFTIADYFSDKTRDVLVYLDADIACKSSIOELIDLN 168
QY 119 LGGNVVGACIDLFVROEG-----YKQI-----GMADGEYFNAGVLLINLKKWRHDI 168

```

Db 147 FAENIAAVV-----ABGELEMTKRSVSLATPGVSG--YFNAGFLINIMPLWTAENI 198
Qy 169 FRKSEWEQKQKDWQ--YQODDILNGLFKGGVCYANSRFPNPTNFAFMANGFASRHTD 226
Db 199 SKKALEMKDEPVVQRTHDDVDLNLVANKARFVDFKFTNQ-----FSLN----- 245
Qy 227 PLYDRTNTAMPVAVS---HYCGSAKPMHRDCTVWGA 260
Db 246 --YELKDSVINPVDAEYFVHIGTPKPMHS---WGA 277

RESULT 8
glycosyl transferase, family 8 SP1770 [imported] - Streptococcus pneumoniae (strain TIGR
C95206
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: C95206
R:Reteljin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-404 <K09>
A:Cross-references: GB:AE005672; PIDN:AAK75844.1; PID:g14973266; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1770

Query Match 12.2%; Score 201.5; DB 2; Length 404;
Best Local Similarity 26.8%; Pred. No. 6.7e-10;
Matches 74; Conservative 48; Mismatches 107; Indels 47; Gaps 12;

Qy 3 IYFAADDNAAVLCVAASVEAHPDTEIRFHVLDAGISEENRAAVALRGCG---NI 58
Db 6 IYFNADNDVLDLETAIKSI--CCYNNCLKFYVENDIDASEFWELMMKRLKTIQSEIVAV 63
Qy 59 RITDVNPEDFAGPLNIRIISTTTARLKGELIADCDVYLIDVDVLYRDLKPLMDTD 118
Db 64 KIVD--HVLKRFHLPLKMLSYATFFRYPIPNFVKE--SRALYLDSDIITGSLDYLFDE 119
Qy 119 LGNNWVGACIDLFVERQEGYKQKIGMADEY---YFNAGVILINLKKWRHDLFKMSCEM 175
Db 120 LQGLAALAEVDSF-----GDVPTNFNSGMLVNVDTWDEDAACSKLLEL 164
Qy 176 VEQYKDVMOYQDOLINGLFKGGVCYANSRFPNPTNFAFMANGFASRHTDPLY-----L 230
Db 165 TNGYHET-AYGQGLIMLNF-----HDMWRKLRDNFNFMGMDSVAHIEGHHKMYEIS 216
Qy 231 DRTNTAMPVAVSHYCGSAKPMHRDCTVWGAERFTEL 266
Db 217 ELKNGDLP-SVIHYTG-VKPMW-----IISNNRREV 246

RESULT 9
C42982
lipopolysaccharide 3-alpha-galactosyltransferase (EC 2.4.1.44) - Escherichia coli (strai
N:Alterate names: lipopolysaccharide core assembly protein rfaI; UDP-D-galactose- (gluc
C:Species: Escherichia coli
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 01-Mar-2002
C:Accession: C42982; S47848; E65163
R:Pradel, E.; Parker, C.T.; Schaitman, C.A.
J. Bacteriol. 174, 4736-4745, 1992
A:Title: Structures of the rfaB, rfaI, rfaD, and rfaS genes of Escherichia coli K-12 and
A:Reference number: A42982; MUID:92325066; PMID:1624461
A:Contents: K-12
A:Accession: C42982
A:Status: preliminary

```

```

A:Molecule type: DNA
A:Residues: 1-339 <PRA>
A:Cross-references: GB:M80599; GB:M86935; NID:g146650; PIDN:AAA24086.1; PID:g146656
A:Note: sequence extracted from NCBI backbone (NCBIN:108180, NCBI:P.108185)
R:Plunkett, G.
submitted to the EMBL Data Library, March 1994
A:Reference number: S47666
A:Accession: S47848
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-339 <PU>
A:Cross-references: EMBL:U00039; NID:g466582; PIDN:AA818604.1; PID:g466765
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Petina, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Sho, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E65163
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-339 <BLAT>
A:Cross-references: GB:AE004440; GB:U00096; NID:g2367252; PIDN:AACT6651.1; PID:g17900
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: rfaI
A:Start codon: GTG
C:Superfamily: lipopolysaccharide glucosyltransferase I
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 11.5%; Score 191; DB 2; Length 339;
Best Local Similarity 21.4%; Pred. No. 4.5e-09;
Matches 67; Conservative 60; Mismatches 142; Indels 44; Gaps 12;

Qy 1 MDIVFAADDNAAVLCVAASVEAHPDTEIRFHVLDAGISEENRAAVALRGCGNIR 59
Db 29 LDIAVGTDKNFLFGGSIASILKYNEGSRCLCFHIFTDFGDDRKYPDALALQYKTRIK 88
Qy 60 FIDVNPEDFAGPLNIRIISTTTARLKGELIADCDVYLIDVDVLYRDLKPLMDTD 118
Db 89 IYLNGDRLSLP-STKMTNATYFRFVYADFIKAPRVLYLDADIIQCGTIEPLNFS 147
Qy 119 LGNNWVGACIDLFVERQEGYKQK-----IGMADGEYFNGAVILINLKKRRHDLFKM 171
Db 148 PDDQVYAMV---TEQADPMWEKRAHSLGVAIGIAG--YFNSGFLININQAQAQVNSAR 202
Qy 172 SCEWEQYQKDV--MOYQDOLINGLFKGGVCYANSRFPNPTNFAFMANGFASRHTDPL 228
Db 203 ATAMLNPEPIIKITHPPQDVLNMLADKLIPADIKYNTQFSILNQLKESFINPYTNDTI 262
Qy 229 YLDRTNTAMPVAVSHYCGSAKPMHRDCTVWG-----AERFTELAGSLTTPPEWNGKLAV 283
Db 263 FI-----HYIGTPKPHD---WAMDYPSQAEMKAKNA-----SPWKNTALL 301
Qy 284 PPTKCMLOKRRKK 296
Db 302 KPNNSNQLRYSAK 314

RESULT 10
T01514
glycosyl transferase homolog T10M13.14 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01514
R:Johnson, A.F.; de la Bastide, M.; Lodhi, M.; Hoffman, J.; Hasegawa, A.; Gnoj, L.; G
Martensen, R.; McCombie, W.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of the Arabidopsis thaliana T10M13 BAC.
A:Reference number: Z14346
A:Accession: T01514
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-346 <JOH>

```

A:Cross-references: EMBL:AF001308; NID:g2104523; PID:g2104536
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 4S
 A:Note: T10M13.14

Query Match 11.5%; Score 190.5; DB 2; Length 346;
 Best Local Similarity 26.3%; Pred. No. 5.1e-09;
 Matches 77; Conservative 36; Mismatches 101; Indels 79; Gaps 15;

Oy 3 IYFAADNVAAYLCVAANSV-EAAHPTEIRFHVLDAGISEENR-----AAVANIRG 54
 Db VAITLDVYELGSIIVANSIIIOHSVCPES-VEFHFI--AVSEETMLESLVSPRPRLK 123
 Oy 55 GGNIFIDVNPEDFAGF-----PLNIRHISITTYARKLGEIADCDKVLDTVL 102
 Db 124 ---FNIFYFAPETVYVGLISSVYQALPEOLN-----YASVYIADLLEPCVKNRYTLD 172
 Oy 103 TDVLVRDGLKPLMDTDLGNWVGA---CIDLFVERQEGYKOKIGMADGEY-----YENA 153
 Db 173 SDLVVVDVADIAKLMKTSLSRIIGAPEYCHANPTKYFTGGFMESEKFSGTFRGRKPCYENT 232
 Oy 154 GVLINLKKRRHRHDFKSCVEVEQYKDYMOYODDILNGL-----FKGVCYANSPRN 207
 Db 233 GVMVIDLKKRRGGYTKRIEKMEIQRERIRYE---LGSPLPFLVSGHVAPISHRMN 288
 Oy 208 FMPTNYAFMANGF-----ASRHTDPLYLDRNTAMPVAVSHYGSAPKPMHR 253
 Db 289 -----QHGLGGDNVRGSGRDLHP-----GVSLLHMSGSGKPMWR 323

RESULT 11

1 lipopolysaccharide 1,2-glucosyltransferase (rfaf) RP476 - Rickettsia prowazekii
 C:Species: Rickettsia prowazekii
 C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C:Accession: A11707
 R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark, U
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A11650; MUID:99039499; PMID:9823893
 A:Accession: A11707
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-517 <AND>
 A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CA14931.1; PID:g386103
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: rfaJ; RP476

Query Match 11.5%; Score 190; DB 2; Length 517;
 Best Local Similarity 22.0%; Pred. No. 9.4e-09;
 Matches 67; Conservative 51; Mismatches 114; Indels 72; Gaps 10;

Oy 1 MDIVEAADNTAAV-LCYAANSVEAAHPDTEIRFHVLDAGISEENRAAVANLRGGNIR 59
 Db 240 LDIALTINDKVAIHAAVYIASLNSDLSDFYRPHI-----VWDSNDPLSQESINKLSMQ 295
 Oy 60 FIDVNPEDFAGFPLNIRHISIT-----TYARKLGEIADCDKVLDTVL 106
 Db 236 YIRDSIDETTFPPDILNKALYKEIKFSNMPSLVMRYLFYDKLFFPNLDSILYLDADIIV 355
 Oy 107 VVDGLKPLMDTDLGNWVGA---IDLFV-----ERQEGYKOKIGMADGEYFNAGVLLNL 160
 Db 356 VLARDNSLKKIDM-NNYIAAGSLDPAITYCIHKVOECKRNVAS-----YKNSGIVFLNL 410
 Oy 161 KWRHRHDFKSCVEVEQYKDYMOYODDILNGLFKGVCYANSPRNPTNYAFMANGF 220
 Db 411 KMRREKQYNNILLETILKNSKCFSPDODLVAVAFONVLYPLSMKWNFC-----459
 Oy 221 ASRHTDPLYLDRNTAMPVAVSHYGSAPKPM-----HRCDC 256
 Db 460 -----TYFEHQSPYFSYFILHY-AEPRPWTEKEELMKTNHDKLDKITQYWRREIT 511

Oy 257 VWGA 260
 Db 512 PWNS 515

RESULT 12

hypothetical protein T14P4.1 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
 C:Accession: C86157
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alon
 Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar,
 anser, N.F.; Hughes, B.; Hultzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: C86157
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-361 <STO>
 A:Cross-references: GB:AF005172; NID:g9972380; PIDN:AG10630.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 11.2%; Score 186; DB 2; Length 361;
 Best Local Similarity 27.6%; Pred. No. 1.3e-08;
 Matches 72; Conservative 32; Mismatches 99; Indels 58; Gaps 11;

Oy 33 FHVLDAGISEENRAAVANLRGGNIREIDVNPEDFAGF-----PLNIRHISIT 81
 Db 114 FHVLSRSSQNLSELRSTFPLTKLIYFAPETVQSLSSVYQALPEOLN-----166
 Oy 82 TYARKLGEIADCDK-VLYLDTVLVYDGLKPLMDTDLGNWVGA---CIDLFVERQEG 137
 Db 167 -YARNYLDLLEPCVKRYIYDSDLVVVDIVYKLMKTGLGQRTIGAPYCHANPTKYFTG 225
 Oy 138 ---YKOKIGMADGE--YFNAGVLLINLKKRRHRHDFKSCVEVEQYKDYMOYODDIL 191
 Db 226 GPMSCKRNGRPFKGNPCYFNTGVAVIDLKKRQRPFRTRIEKMEIQRERIRYE-----L 281
 Oy 192 NGL-----FKGVCYANSPRNPTNYAFMANGFASRHTDPLYLDRNTAMPVAVSHYC 245
 Db 282 GSLPFLVYFAGHVAPISHRMNQHGLGGDNVRGSGRDLHSG-----PVSLLHMS 330
 Oy 246 GSAKPMHR-----DC---TW 258
 Db 331 GSGKPMRLRDSKRLPCPLDTLW 351

RESULT 13

hypothetical protein T20K14_80 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
 C:Accession: T51525
 R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.;
 submitted to the Protein Sequence Database, August 2000
 A:Reference number: Z25394
 A:Accession: T51525
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-572 <SAT>
 A:Cross-references: EMBL:AL911143
 A:Experimental source: cultivar Columbia; BAC clone T20K14
 C:Genetics:
 A:Map position: 5

A;Introns: 43/2; 114/1; 129/3; 181/3; 378/3; 479/3
A;Note: T20K14_80

Query Match	11.1%;	Score 184;	DB 2;	Length 572;
Best Local Similarity	23.5%;	Pred. No. 3.6e 08;		
Matches 73;	Conservative 40;	Mismatches 108;	Indels 90;	Gaps 11.

```

QY      9  DNVAAYUJCVAKSVSEAAHPDELEIFEHYLD-----AG----- 39
      || : : : : : || : : : : : ||
Db      254 DNIIAAVSVMSSAAVOSSSKREKIVFHIITIDDKITYAGAHSMWFALNSVAPALVEYKGVHGFQD 313
      || : : : : : || : : : : : ||
QY      40  -ISEBNRAVA-----NLRGGINREIDVNPEDPAGPRLNR---HISIT 82
      || : : : : : || : : : : : ||
Db      314 WLIRENPVLEAEVESHGVRDYHYHGNHVAAGANL--TETTPRTFAS-KIOSRSKPYISILN 370
      || : : : : : || : : : : : ||
QY      83  YARKLGEIYIDCKVYLIDPVLVYRGKILPYMOTDLGGMMGVGACIDLEPERQEGYKOKI 142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      371 HIRIYIEPLFNIDKVVFLDDDIYVGQDLIRPLMDVLDGKGVKNGA-----VETCRGEDEWY 425
      || : : : : : || : : : : : ||
QY      143 GMADEGEYFN-----AGVLLINLKKWRHIDFKMSCEW--EQRKYDM 183
      || : : : : : || : : : : : ||
Db      426 MSKRLRNYPFNFSHPDLAKHLDEPCAAWAGYGMNIFEDLOAMERKTNIRETYHSLMRENLKSNL 485
      || : : : : : || : : : : : ||
QY      184 QYQODIDLNG---LEKGGVCCANREFMFMTNVAFPMANGASRTIDLYLIDRNTIAPVA 240
      || : : : : : || : : : : : ||
Db      486 TMMKLGITPLPALLINFKSHVHITIDSSWMLGLGYOSTKN-----IENYKKA 511

```

RESULT 14

lipopolysaccharide glucosyltransferase I (EC 2.4.1.58) - *Escherichia coli* (strain K-12)

C:Species: Escherichia coli
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 01-Mar-2002
C:Accession: S47847; D65163; D42982

R;Plunkett, G.
submitted to the EMBL Data Library, March 1994
A;Reference number: S47666

A:Accession: S47847
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <PLU>
A:Cross-references: EMBL:U00039, NID:9466582, PIDN:AAB18603.1, PID:9912479
R:Blatter, F.R., Plunkett II, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Cai,
A., Rose, D.J., Mau, B., Sha, Y.

Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: D65163
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-338 <BLAT>
 A:Cross-references: GB:AE000440; GB:U00096; NID:g9267252; PIDN:AACT6650.1; PID:g1790056
 A:Experimental source: strain K-12, substrain MG1655
 A:Pradel, E.; Parker, C.T.; Schnaitman, C.A.
 B:Microbiol. 174, 4736-4745, 1992

Query Match	10.9%;	Score 180;	DB 2;	Length 338;
Best Local Similarity	23.7%;	Pred. No. 4.2e-08;		
Matches 64;	Conservative 48;	Mismatches 118;	Indels 40;	Gaps 9;

```

QY 1 MDIYFAADNDNAAIYICVAAKSVYEAHAHPDTEIRFHYLD-----AGISEEKRAAVA 49
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 28 LNVAVGVANVYALDGVGVASITSIVLNNRHINIDFYIADVYNDGFQOKIAKLAEQNOL-- 84
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 50 ANLRGGGNIREFIDVPEEFAGFPLN---IRHISITTYARLKGEIADCDKVLKLDPTVL 106
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 -----RITLYIRINTDKLOCLPCTQVWSRAMYPRFLFAFOLLG---LTLDRLYLVDADV 134
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 107 VRDGLKPLMPDDLDGNNWGACIDLFVERQEGYKOKIGADE---YYNNAVLLINLKKV 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 CKGDISQILLHLGLNKA-VAAYVKKDVEPMOE---KAVSRISDELLGOYFNSSGVVYLDIKKW 191
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 RRHDIFFKMSCEVEQYKRDVMOYODDILNGELFGGVGCYANSRFEMPRNPAVFMAANGFASR 223
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 ADAKLTAKALSIILMSKDNVYKPRDQDVANVLLIKGWTFLPREVN---IYYITKSLKOKT 248
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 HTDPLYLDRNTAMPVAVSHYCGSAKPPMR 253
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 HONYKLLTESTLL---IHYTGATKPPMK 274
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15
A95206

glycosyl transferase, family 8 SPI767 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae

C:\Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
R:\Accession: A95206
R:\Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H.
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple
nson, T.; Hickey, E.K.; Holt, I.E.

A: Authors: Lofthus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris A.; Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.
A: Reference number: A95000; MUID: 21357209; PMID: 11463916

A:Accession: A95206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-814 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75842.1; PID:q14973264; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:

Query Match	10.8%	Score 179.5	DB 2	Length 814
Best Local Similarity	28.3%	Fred. No. 1.4e-07		
Matches 79; Conservative	38;	Mismatches 109;	Indels 53;	Gaps 16

[illegible]

Job time : 15.5805 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 7.44739 Seconds
(without alignments)
1704.189 Million cell updates/sec

Title: US-10-007-267-4
Perfect score: 1657
Sequence: 1 MDIVFADNDYAAVLCVAAK.....KCMQRMWRKKLSARLEKRY 306

Scoring table: BIOSOM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	536	32.3	330	1 Y258_HAEIN	P43974 bacillus
2	252	15.2	286	1 GSPA_BACSU	P25148 bacillus su
3	219	13.2	337	1 RPAI_SALTY	P19816 salmonella
4	191	11.5	339	1 RPAI_ECOLI	P27128 escherichia
5	180	10.9	338	1 RPAI_ECOLI	P27129 escherichia
6	173	10.4	336	1 RPAI_SALTY	P19817 salmonella
7	110.5	6.7	1548	1 UGGG_DROME	P09332 drosophila
8	103.5	6.2	1448	1 UGGG_SCHPO	P09140 schizosacch
9	94.5	5.7	393	1 MN10_YEAST	P50108 saccharomyc
10	93	5.6	1236	1 TOP2_LEICH	P61078 leishmania
11	89	5.4	505	1 SCRY_SALTY	P22340 salmonella
12	88	5.3	283	1 V635_MERJA	P58052 methanococc
13	87	5.3	394	1 OM1B_CHLPR	P23421 chlamydia t
14	86.5	5.2	1202	1 DPOM_ASCIM	P22374 asaccharid
15	85	5.1	576	1 GRK6_MOUSE	P070293 mus musculu
16	84	5.1	804	1 VP5_MTV	P13266 wound tumor
17	83.5	5.0	332	1 GRK6_RAT	P97111 ratius norv
18	83	5.0	576	1 LAGE_HUMAN	P09461 homo sapien
19	82.5	5.0	756	1 LAGE_MOUSE	P09461 homo sapien
20	82.5	5.0	756	1 LAGE_MOUSE	P09461 homo sapien
21	81.5	4.9	404	1 GIGC_MYCTU	P05314 mycobacteri
22	81.5	4.9	464	1 RFW7_ECOLI	P37741 escherichia
23	81	4.9	187	1 PYRE_METAC	P58859 methanocarc
24	81	4.9	389	1 OM1N_CHLPR	P07430 chlamydia p
25	81	4.9	389	1 OM1N_CHLPR	P07430 chlamydia p
26	81	4.9	1113	1 MYSD_DICDI	P24105 chlamydia p
27	80.5	4.9	219	1 OSEB_ECOS7	P43974 bacillus
28	80.5	4.9	219	1 OSEB_ECOS7	P43974 bacillus
29	80	4.8	900	1 GIG1_YEAST	P43974 bacillus
30	79.5	4.8	295	1 YF67_ARCFU	P43974 bacillus
31	79.5	4.8	332	1 GUYG_MOUSE	P08705 archaeoglob
32	79.5	4.8	332	1 GUYG_MOUSE	P08705 archaeoglob
33	79	4.8	333	1 OM1K_CHLPR	P08705 archaeoglob

34	79	4.8	437	1 PEPG_LACDL	P94869 lactobacilli
35	79	4.8	557	1 PGMU_EMENI	O99931 emericella
36	79	4.8	1411	1 YK63_CABEL	P34342 caenorhabdi
37	78.5	4.7	1019	1 IDE_RAT	P35559 ratius norv
38	78	4.7	422	1 LAGE_HUMAN	P29822 agrobacteri
39	78	4.7	1694	1 IGAO_HAEIN	P44969 haemophilus
40	78	4.7	1702	1 IGA2_HAEIN	P45384 haemophilus
41	77.5	4.7	99	1 RS24_PYRAB	O99y20 pyrococcus
42	77.5	4.7	219	1 OSEB_SALTY	O8aeq3 salmonella
43	77.5	4.7	305	1 MIAA_THEMA	O9wy25 thermotoga
44	77.5	4.7	461	1 SYC_SALTY	O8z8p6 salmonella
45	77.5	4.7	461	1 SYC_SALTY	O8z8p6 salmonella

ALIGNMENTS

```

RESULT 1
ID      Y258_HAEIN      STANDARD:      PRT:      330 AA.
AC      P43974: P44597; P71347;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical protein HI0258.
GN      HI0258.
OS      Haemophilus influenzae.
OC      Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC      Haemophilus.
OX      NCBI_TaxID=727;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=RD / KW20 / ATCC 51907;
RX      MEDLINE=95350630; PubMed=7542800;
RA      Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA      Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA      McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA      Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA      Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA      Uterback T.R., Hanna M.C., Nguyen B.T., Saudek D.M., Brandon R.C.,
RA      Fine L.D., Fritchman J.L., Fuhrmann J.L., Georgagen N.S.M.,
RA      Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA      Venter J.C.;
RT      "Whole-genome random sequencing and assembly of Haemophilus influenzae
RL      Rd.";
RN      Science 269:496-512(1995).
RP      [2]
RC      REVISIONS.
RA      White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.;
RL      submitted (SRP-1996) to the EMBL/Genbank/DBJ databases.
CC      -I- SIMILARITY: TO N.GONORRHOEA LGT AND WEAK. TO B.SUBTILIS GSPA.
CC      LIPOPOLYSACCHARIDE 1,3-GALACTOSYLTRANSFERASE (RPA1) AND TO
CC      LIPOPOLYSACCHARIDE 1,2-GLUCOSYLTRANSFERASE (RPAJ).
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL: U32711; AAC21924.1; -.
DR      TIGR; HI0258; -.
DR      InterPro: IPR002495; GT_8.
DR      Pfam: PF01501; Glyco-transf_8; 1.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 330 AA; 39073 MW; CSA6CA0207F24F9A CMC64;
Query Match 32.3%; Score 536; DB 1; Length 330;
Best Local Similarity 36.9%; Pred. No. 1,1e-40;
Matches 113; Conservative 58; Mismatches 109; Indels 26; Gaps 8;

```

[illegible]

RESULT 2				
GSPL_BACSU	ID	GSPL_BACSU	STANDARD;	PRT; 286 AA.
AC	P25148;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	General stress protein A.			
CN	GSPL OR IPA-12D.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RC	MEDLINE=92216127; PubMed=1806041;			
RA	Glaser P., Kunst F., Debarbouille M., Vertes A., Danchin A.,			
RA	Deonder R.;			
RT	"A gene encoding a tyrosine tRNA synthetase is located near sacs in			
RT	Bacillus subtilis.";			
RL	DNA Seq. 1:251-261(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RC	MEDLINE=95020537; PubMed=7934828;			
RA	Glaser P., Kunst F., Arnold M., Coudart M.P., Gonzales W.,			
RA	Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,			
RA	Presseau E., Santana M., Schneider E., Schweizer J., Vertes A.,			
RA	Rapoport G., Danchin A.;			
RT	"Bacillus subtilis genome project: cloning and sequencing of the 97			
RT	kb region from 325 degrees to 333 degrees.";			
RL	Mol. Microbiol. 10:371-384(1993).			
RN	[3]			
RP	SEQUENCE OF 110-286 FROM N.A.			
RC	STRAIN=QB1072;			
RC	MEDLINE=90337338; PubMed=2116367;			
RA	Zukowski M.M., Miller L., Cosgwell P., Chen K., Aymerich S.,			
RA	Steinmetz M.;			
RT	"Nucleotide sequence of the sacs locus of Bacillus subtilis reveals			
RT	the presence of two regulatory genes.";			
RL	Gene 90:153-155(1990).			
RN	[4]			
RP	CHARACTERIZATION, AND SEQUENCE OF 1-28.			
RC	MEDLINE=95286551; PubMed=7768864;			
RA	Antelmann H., Bernhardt J., Schmid R., Hecker M.,			
RA	"A gene at 333 degrees on the Bacillus subtilis chromosome encodes			
RT	the newly identified sigma B-dependent general stress protein GspA.			
RL	J. Bacteriol. 177:3540-3545(1995).			

```

CC      -I- INDUCTION BY DIFFERENT STRESSES SUCH AS HEAT SHOCK AND SALT
CC      STRESS AND BY STARVATION
CC      -I- SIMILARITY: TO H.INFLUENZAE H10259.
CC      -I- SIMILARITY: TO LIPIDOLYSACCHARIDE 1,3-GALACTOSYLTRANSFERASE (REAI)
CC      AND TO LIPIDOLYSACCHARIDE 1,2-GLUCOSYLTRANSFERASE (REAF).
CC
CC      ---
CC      THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC      BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC      THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC      USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC      MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC      ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/
CC      OR SEND AN EMAIL TO license@isb-sib.ch).
CC
CC      -----
DR      EMBL: X52480; CAA36721.1; ALT_INT.
DR      EMBL: X73124; CAA51568.1; -
DR      EMBL: M29333; AAA75337.1; -
DR      EMBL: Z99123; CAB15869.1; -
DR      PIR: S16423; S16423.
DR      Subtilist; BG10558; gspDA.
DR      InterPro: IPR002495; GL_8.
DR      Pfam: PF01501; glyco_transf_8; 1.
KM      Complete proteome.
FT      CONFLICT 12 12 C -> A (IN REF. 4).
FT      CONFLICT 27 27 S -> V (IN REF. 4).
SQ      SEQUENCE 286 AA; 33522 MW; A282B6E3B6518DE CRC64;

```

```

Query Match          15.28; Score 252; DB 1: Length 286;
Best Local Similarity 26.48; Pred. No. 2,6e-15;
Matches 69; Conservative 54; Mismatches 118; Indels 20; Gaps 9;

OY 1 MDIYVAADDNNAAYL-CYAAKSVEAAHDDTEIRFVLDLAGISEENRAAV-AANLRGGGNI 58
   | | | | | | | | : | : | : | : | : | : | : | : | : | : | : |
DB 7 MHIVSCADDNARHNGMFVSLTMMDDEREVKLYIDGGIKPKDKKRLDEETTLKFGVPI 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 59 RFIDVNPDEFGAFPIINRHSITTYARLKLEGIAD--CDKVLVLDTVLVDGKLKPLMD 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 67 EFLVEDTNMYE-HAVESHITKAAYRIRISIDPLIDKESIKRMVYIDCALVLEDISKLMD 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 117 TDLGNWNGACIDLFEVEROEGYKOKIGMADGEVYFNAGVLLINLKKMRBIRFKKSCMV 176
   | | | | | | | | : : : : | | | | : : : : : : : : : : : : : :
DB 126 LDIAPYTAAYEED--AGQHERLKEMANTDTGKFFNSGIMIIDESWKKQNTTEKVINFI 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 177 EGY--KDVMOYODODILNGLFKGGCYANSRFENFMPTVYAFMANGFASRHTDPLXDR-- 232
   : : : : | | | | : : : : : : : : : : : : : : : : : : : :
DB 183 NENRPEDPLVLDQDALNALIYLDQGYELHPRNN--AQYTYML-----KLKTPSTLGRKQ 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 233 -TNTAMPYAVSHYCGSAKPMH 252
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 236 YNETRENPAIVHFCGGEKPMN 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
RFAL_SALT
ID RFAL_SALT STANDARD; PRT; 337 AA.
AC P19816; O68267;
DT 01-FEB-1991 (Rel. 17, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipopolysaccharide 1,3-galactosyltransferase (EC 2.4.1.44).
GN RFAL OR WAI1 OR STM3718.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RX MEDLINE=98204873; PubMed=9535865;
RA "The assembly system for the lipopolysaccharide R2 core-type of
RT Escherichia coli is a hybrid of those found in Escherichia coli K-12
RT and Salmonella enterica. Structure and function of the R2 weak and
```



```
Db 89 IYLNGDRLSLP-STKNMTHAIFRFVIADYFINKAPKVLDDADIICOGTEPLINFS 147
QY 119 LGSNWVACIDLEVEREGYKQK-----IGMADGEYFNAGVLLINLKKRRHDFKM 171
Db 148 FPDQKVAWVY---TEGADWMEKRAHSLGVAIGAKG--YNSGFLILINTQMMAQOVSAR 202
QY 172 SCWEVQYKDY--MOYODODILNGLFKGVGYCANSRPNF-MPTNYAFMANGFASRHTDPL 228
Db 203 AIALMNEPEIKRITHPDODVNLNMLADKLIFADIKYNTQFSINYQLKESFINPVTNDTI 262
QY 229 YLDRNTNAMPVAVSHYCSGSAKPMHRDCTWNG-----AERTELAGSLTTYPEERKGLAV 283
Db 263 FI-----HITGPTKPMHD---WMDYPVSOAFMEAKNA-----SPWKNTALL 301
QY 284 PPTKCMLOMRKK 296
Db 302 KPNNNSQLRYSAK 314

RESULT 5
RFAL_ECOLI STANDARD; PRT; 338 AA.
ID RFAL_ECOLI
AC P27129;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipopolysaccharide 1,2-glucosyltransferase (EC 2.4.1.58).
GN RFAL OR WAAJ OR B3626.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_Taxid=562;
RN 11]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92325066; PubMed=1624461;
RA Pradel E., Parker C.T., Schnaitman C.A.;
RT Structures of the rfaB, rfaI, rfaJ, and rfaS genes of Escherichia
RT coli K-12 and their roles in assembly of the lipopolysaccharide
RT core-";
RL J. Bacteriol. 174:4736-4745(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -1- CATALYTIC ACTIVITY: UDP-glucose + lipopolysaccharide = UDP + D-
CC glucosyl-lipopolysaccharide.
CC -1- PATHWAY: Lipopolysaccharide core biosynthesis.
CC -1- SIMILARITY: TO S.TYPHIMURIUM RFAL AND TO RFAL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M80599; AAA24087.1; -
CC DR EMBL; U00039; AAB18603.1; -
CC DR EMBL; AE000440; AAC76650.1; -
CC DR PIR; D42982; D42982.
CC DR EcoGene; EG11353; rfaJ.
CC DR InterPro; IPR002495; GT_8.
CC DR Pfam; PF01501; Glyco_transf_8; 1.
CC DR Lipopolysaccharide biosynthesis; Glycosyltransferase; Transferase;
CC complete proteome.
CC KW
CC CONFLICT 310 338 RYKHLVQHHYISGIAGVCYLCKRYRK -> DINITF
CC FT (IN REF. 1).
```

```
SQ SEQUENCE 338 AA: 39040 MW: 819428EA13F1959A CRC64;
Query Match 10.9%; Score 180; DB 1; Length 338;
Best local Similarity 23.7%; Pred. No. 9,1e-09;
Matches 64; Conservative 48; Mismatches 118; Indels 40; Gaps 9;
QY 1 MDIVFADNDVYAAVLCVAAKSVEAHPDEIRPHVID-----AGISEENRAVA 49
Db 28 LNVAVGDVANYLDVGVSITSIVLNNRHINLDFYIADVYNDGFPQIAKLARQNL--- 84
QY 50 ANLRGGNIFFDIVNPEDFGFLN---IRHISITTAARKLGEIYADCKVLYLDTVL 106
Db 85 -----RITLYRINTDKLOCLPCTOWSRAMVRLFAFQLG---LTLDRLLYLDADV 134
QY 107 VRDGLKPLMDTDLGSNMGVACIDLEVEREGYKOKIGMADGE---YFNAGVLLINLKKV 163
Db 135 CKGDISQLHLGLNGA-VAAVVNDVPEMGE--KAVSLSPPELLGGYFNSGVYLLDKW 191
QY 164 RRHDFKMSCEWVEQYKDYMOYODODILNGLFKGVGYCANSRPNFMTNYAFMANGFASR 223
Db 192 ADAKLFKALSIILMSKDNVYKYPDQVMANVLLGMFLFLEPREYN---TIYIKSELKDT 248
QY 224 HTDPLVLDRTNTAMPVAVSHYCSGSAKPMHR 253
Db 249 HQNYKKLITESTLL---IHYGTATKPMWK 274

RESULT 6
RFAL_SALTY STANDARD; PRT; 336 AA.
ID RFAL_SALTY
AC P19817; 068266;
DT 01-FEB-1991 (Rel. 17, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipopolysaccharide 1,2-glucosyltransferase (EC 2.4.1.58).
GN RFAL OR WAAJ OR STM3717.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_Taxid=602;
RN 11]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=98204873; PubMed=9535865;
RA Heinrichs D.E., Monteiro M.A., Perry M.B., Whitfield C.;
RT "The assembly system for the lipopolysaccharide R2 core-type of
RT Escherichia coli is a hybrid of those found in Escherichia coli K-12
RT and Salmonella enterica. Structure and function of the R2 waak and
RT waal homologs."
RL J. Biol. Chem. 273:8849-8859(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGGC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE OF 1-241 FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=91045080; PubMed=2255496;
RA Carstenius P., Flock J.-I., Lindberg A.;
RT "Nucleotide sequence of rfaI and rfaJ genes encoding
RT lipopolysaccharide glycosyl transferases from Salmonella
RT typhimurium."
RL Nucleic Acids Res. 18:6128-6128(1990).
RN [4]
RP SEQUENCE OF 242-336 FROM N.A.
```


RA Oliver K., Harris D., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
 CC -!- FUNCTION: PUTATIVE GALACTOSYLTRANSFERASE REQUIRED FOR POLARIZED
 CC GROWTH AND EFFICIENT BUDDING.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (potential).
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 34.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: LA2540; AAB48372.1; -
 CC DR EMBL: U31446; AAC49280.1; -
 CC DR EMBL: 249701; CA89731.1; -
 CC DR SGD: S0002853; MN10.
 CC Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KM Golgi stack.
 FT DOMAIN 1 52 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 53 73 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 74 393 LUMENAL (POTENTIAL).
 FT CONFLICT 238 238 N -> I (IN REF. 2).
 SQ SEQUENCE 393 AA; 46748 MW; 5FA7DBE3122EEC8E CRC64;

Query Match 5.7%; Score 94.5; DB 1; Length 393;
 Best Local Similarity 22.4%; Pred. No. 0.5;
 Matches 67; Conservative 38; Mismatches 89; Indels 105; Gaps 17;

QY 48 VAANLGGNINRIFIDNPDA-----GEPINIRHISTTYARLKGEY 91
 DB 126 LAAN-BGGVGLW--KNEBEMAEIGSINAKAYAKRHGAYLTIK--DLTTSKRY-HEX 179
 QY 92 IADCDKV-----LYLDTVLVR-----DGLKPLMDTLGG-NW 123
 DB 180 REGMOKVDILROTFRFPNAEFMWMLDITMIEPSKLEEHIFDLFTLADRELKSFNP 239
 QY 124 VGACIDL-EVEROEGKKOIGMADGEYFNAGVLLINKKRRHDFKSCEN---VQG 178
 DB 240 LNRDDIPYVDSEMEFLITQDCG--FNLSFLLKNSWS--LLDDMMWDVPLYEQ 294
 QY 179 YDVMQYODDILNGLFKGCVCYANSRFPNPNVAFMANGFASRHTDPLYDRTTAMP 238
 DB 295 KIMWEHRQDALLALEYENP-WISRIGFLP-----LRTINAFPP 334
 QY 239 VAVSHYGSAGKPMH-----RDCYVGAERFTELAGSLTTPPEWRGK 280
 DB 335 GACSEYSGDSRFYSEKDHDFVVMAGCNFRGC--WGEMQY-----YTTLMEXLNK 365

RESULT 10
 TOP2_LEICH STANDARD; PRT; 1236 AA.
 ID TOP2_LEICH
 AC 061078;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA topoisomerase II (EC 5.99.1.3).
 GN TOP2.
 OS Leishmania chagasi.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=44271;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HOM/BR/86/L669;
 RA Tepe-Lansdel T., Mann B.J., Labombard M., Macdonald T., Slunt K.M.,
 RA Pearson R.D.;
 RT "Isolation of a gene encoding a DNA topoisomerase II of Leishmania
 RT (Leishmania) chagasi.";
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.

CC -!- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 CC MAKES DOUBLE-STRAND BREAKS.
 CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
 CC of double-stranded DNA.
 CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMS
 CC RELAX ONLY NEGATIVE SUPERCOILS.
 CC -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: AF051307; AAC05295.2; -
 CC DR HSSP: P06786; IBBW.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR001241; DNA_topoisoi.
 DR InterPro: IPR002205; DNA_topoisoi.
 DR Pfam: PF00204; DNA_gyraseb; 1.
 DR Pfam: PF00521; DNA_topoisoi; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR PRINTS: PR00418; TP12FAMILY.
 DR PRODOM: PD000616; DNA_topoisoi; 1.
 DR PRODOM: PD000742; DNA_topoisoi; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00433; TOP2C; 1.
 DR SMART: SM00434; TOP4C; 1.
 DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
 DR Isomerase; Topoisomerase; ATP-binding.
 FT NP_BIND 137 142 ATP (POTENTIAL).
 FT ACT_SITE 775 775 DNA_CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 1236 AA; 138968 MW; A3209B95A078045C CRC64;

Query Match 5.6%; Score 93; DB 1; Length 1236;
 Best Local Similarity 21.0%; Pred. No. 2.8;
 Matches 64; Conservative 36; Mismatches 111; Indels 94; Gaps 13;

QY 15 LCVAAKSVDAHPDEIRRHVID-----AGISEERAFAVANL-----RGGN 57
 DB 819 LCGAGVGICFGAFTTPSHPLDVSAVRAKINGESAKOVRLVPMVAVFGGTVARGPD 878
 QY 58 IRFIDVNPEDFAGFPLNIRHIS-----ITTYARLKGEYIADCDKLYLDPT 103
 DB 879 NEFIANG--KYAHAPRGRRHISIPMTSIEAFRLHSSLSADYQRIADYSGANHIDI 936
 QY 104 DYLVNRDGLKPLW---DTDLGMMWVACIDLFEVEROEGYKOKIGMADGEYFNAGVL---- 156
 DB 937 DLIVRDGSLTTWAECTD-----ALAQRI-YINGTVFSPTGLSPID 978
 QY 157 --LINKKW---RRHDIK-----MSCWEVEQK---DYMQYQDD 189
 DB 979 SDSLPLYQNHVDRLLDYLRRRTKRGKLMKMDLARIQSTRKEVHNHROGQIDFLNATDPT 1038
 QY 190 IINGLFKGCVCYANSRFPNM---PTNYAFMANGFASRHTDPLYDRTNAMPVAVSHYGC 246
 DB 1039 LHKTKVGLLVNRDESFDYILKKPIF-----FTFTSTENKIQADIAKIQADIEELKRTT 1092
 QY 247 SAKPW 251
 DB 1093 PVKMW 1097

RESULT 11
 ID SCRY_SALT
 AC SCRY_SALT
 P22340;
 STANDARD; PRT; 505 AA.

```

DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Sucrose porin precursor.
GN SCRY.
OS Salmoneella typhimurium.
OC Bacteriia; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneella.
OC NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-33.
RC STRAIN=6153-62;
RX MEDLINE=91100329; PubMed=1846143;
RA Hãrdesty C., Ferran C., Dirlenzo J.M.;
RT "Plasmid-mediated sucrose metabolism in Escherichia coli:
RT characterization of scry, the structural gene for a
RT phosphoenolpyruvate-dependent sucrose phosphotransferase system
RT outer membrane porin.";
RL J. Bacteriol. 173:449-456(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91312133; PubMed=1649946;
RA Schmid K., Ebner R., Jahreis K., Tengeler J.W., Tilgemyer F.;
RT "A sugar-specific porin, Scry, is involved in sucrose uptake in
RT enteric bacteria.";
RL Mol. Microbiol. 5:941-950(1991).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ÅNGSTRÖMS).
RX MEDLINE=98100081; PubMed=9437428;
RA Forst D., Wette W., Wacker T., Diederichs K.;
RT "Structure of the sucrose-specific porin Scry from Salmoneella
RT typhimurium and its complex with sucrose.";
RL Nat. Struct. Biol. 5:37-46(1998).
CC -1- FUNCTION: Porin for sucrose uptake.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- DOMAIN: THE C-TERMINUS HELPS TO ANCHOR THE PORIN TO THE OUTER
CC MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE LAMB FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M38416; AAA98417.1; -
DR EMBL; X57400; CAA40656.1; -
DR PIR; S15193; S15193.
DR PIR; A39127; A39127.
DR PDB; 1A0S; 10-JUN-98.
DR PDB; 1A0T; 18-MAR-98.
DR InterPro; IPR003192; Porin_Lamb.
DR Pfam; PF02264; Lamb; 1.
KW Transport; Sugar transport; Outer membrane; Transmembrane; Porin;
KW Signal; Plasmid; 3D-structure.
FT SIGNAL 1 22
FT CHAIN 23 505 SUCROSE PORIN.
SQ SEQUENCE 505 AA; 55467 MW; DD37733EC22A2135 CRC64;
Query Match 5.4%; Score 89; DB 1; Length 505;
Best Local Similarity 19.3%; Pred. No. 2.1;
Matches 50; Conservative 35; Mismatches 82; Indels 92; Gaps 10;
QY 13 AYCAVAAKSEVAHPDTEI-----RFHVLDAGISE-ENRAAVANLRGGNIRF----- 60
DB 8 AMLIALLTSASASHAOTDISTIFARLNALEKRIQEAENRQOTAEENRGAAEKKVQQLTAQ 67
QY 61 -----IDVNPEDFAGFPLINRHSITTYTARLKLEGYIADCDKVLVLDPT 103

```

```

DB 68 OOKNONSSTOEVAQRTARLEKKADKSGFEFH-----GYAR-----SGVIMNDS 110
QY 104 DVLVRG--LKPMDPDLGNWNGACIDLFVERQEGYQ-----KIGADEEYF 151
DB 111 GASTKSGAVITPAGETGAIQRLGNQADYVERMLEKHQTLONGATVTRVWVADQTSI 170
QY 152 NAGVLLINLKRRHRHDIKFKSCWEVEQYKDV-----MOYQODIINGLFKGVCAVANSR 205
DB 171 N-----DWTASTSDLVNRQAFVELGNLPTFAGPRKSGSTLMAGKR 209
QY 206 F-----NEMPTNYAFMA 217
DB 210 FDRDNFDIHWISDVVELA 228
RESULT 12
Y635_METUA STANDARD; PRT; 283 AA.
ID Y635_METUA
AC 058052;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Hypothetical protein MJ0635.
GN MJ0635.
OS Methanococcus jannaschlii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_Taxid=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RC MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glödek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Utherback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Kleen H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschlii.";
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67511; AAB98633.1; -
DR TIGR; M0635; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 283 AA; 32502 MW; ICE52FC8457C5E20 CRC64;
Query Match 5.3%; Score 88; DB 1; Length 283;
Best Local Similarity 22.1%; Pred. No. 1.3;
Matches 50; Conservative 47; Mismatches 65; Indels 64; Gaps 14;
QY 10 NYAAVLCVAAKSEVAHPDTEIRFHVLD-----GISENRAAVAN-----LRGGNIRF 60
DB 43 DYNGEYVNFKNIEDSKYSEYRSKILNALRLGNANGKGTGIYINTKRIRFSGSGYDRF 102
QY 61 -----IDVNP-EDFAGFPL-----NIRHSITTY-----ARLKGEY 91
DB 103 VITIEGDYDFDKFNKLKEIGVNPVEYGFKLYTRPNDDKIALTFYKDWIINGTQGYV 162
QY 92 IADCDKVLVLDVLYRD-GLKPLMD---TDLGGNNGACI-DLFVERQEGYQKIGMAD 146
DB 163 --DCINVINQEMDSLKNREVMETIYRLPSD-----ACVEVSGTSPSPWKT---VAE 210

```


GRK6_MOUSE STANDARD: PRT: 576 AA.

ID GRK6_MOUSE 070294: 070294: 30-MAY-2000 (rel. 39, Created)

AC 30-MAY-2000 (rel. 39, Last sequence update)

DT 30-MAY-2000 (rel. 41, Last annotation update)

DT 15-JUN-2002 (rel. 41, Last annotation update)

DE G protein-coupled receptor kinase GRK6 (EC 2.7.1.1) (G-protein-coupled receptor kinase 6).

GN GRK6 OR GRK6.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS GRK6A AND GRK6B).

RA Premont R.T., Lefkowitz R.J.;

RT "Genomic organization of the GRK5 and GRK6 genes: conserved structure among the GRK subfamily of G protein-coupled receptor kinases and evolution of the GRK gene family."

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS GRK6A AND GRK6B).

RA STRAIN=129/SVJ, and C57BL/6J; TISSUE=Thymus;

RC Moepers B., Vatter P., Frode R., Maechler F., Gierschik P.;

RT "Primary structure of murine G-protein-coupled receptor kinase 6 splice variants predict differential regulation by posttranslational modifications."

RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: SPECIFICALLY PHOSPHORYLATES THE ACTIVATED FORMS OF G PROTEIN-COUPLED RECEPTORS (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Membrane-bound.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; GRK6A (SHOWN HERE) AND GRK6B; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. GPCR SUBFAMILY.

CC -1- SIMILARITY: CONTAINS 1 RGS DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

DR EMBL: AF040747; AAC09268.1; -

DR EMBL: AF040748; AAC09269.1; -

DR EMBL: AF040749; AAC09270.1; -

DR EMBL: Y17967; CAA76975.1; -

DR EMBL: Y17967; CAA76975.1; -

DR EMBL: Y17967; CAA76975.1; -

DR EMBL: Y15798; CAA75789.1; -

DR EMBL: Y15799; CAA75790.1; -

DR HSSP: Q63450; 1A06.

DR MGD: MGI:1347078; Gprk6.

DR InterPro: IPR000719; Euk_pkinase.

DR InterPro: IPR000239; GPCR_kinase.

DR InterPro: IPR000961; Pkinase.C.

DR InterPro: IPR000342; Regl_Gprotein.

DR InterPro: IPR002290; Ser_thr_pkinase.

DR Pfam: PF00069; pkinase.1.

DR Pfam: PF00615; RGS.1.

DR PRINTS: PR00717; GPCR_KINASE.

DR ProDom: PD000001; Euk_pkinase.1.

DR SMART: SM00315; RGS.1.

DR SMART: SM00133; S_TK_X.1.

DR SMART: SM00220; S_TKc.1.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.

DR PROSITE: PS00111; PROTEIN_KINASE_DOM.1.

DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.

DR PROSITE: PS00132; RGS.1.

KW Transferase; Serine/threonine-protein kinase; ATP-binding; Phosphorylation; Lipoprotein; Palmitate; Alternative splicing.

FT DOMAIN 1 185 N-TERMINAL.

FT	DOMAIN	186	448	PROTEIN KINASE.
FT	DOMAIN	449	576	C-TERMINAL.
FT	DOMAIN	53	171	RGS.
FT	NP_BIND	192	200	ATP (BY SIMILARITY).
FT	BINDING	215	215	ATP (BY SIMILARITY).
FT	ACT_SITE	311	311	BY SIMILARITY.
FT	MOD_RES	484	484	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	485	485	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	LIPID	561	561	PALMITATE (BY SIMILARITY).
FT	LIPID	562	562	PALMITATE (BY SIMILARITY).
FT	LIPID	565	565	PALMITATE (BY SIMILARITY).
FT	VARSPLIC	560	576	DCCGNCSDSEELPTRL -> RIAVGATATVRKSSPPASSP QAEAPTGGWR (IN ISOFORM GRK6B).
FT	SEQUENCE	576 AA:	65978 MW;	91EC2028CAFA91 CRC64;

Query Match 5.1%; Score 85; DB 1; Length 576;

Best Local Similarity 21.3%; Pred. No. 5.7; Mismatches 121; Indels 96; Gaps 18;

Matches 71; Conservative 46;

32 REFVLDAGISEENRAA--VAANLRGGGNIREFIDVNPEDPAGFPLNIRHSITTYARLKLG 89

244 RF-VVSLAAYETKDALCLVLTLMNGDLKF-----HIYHGAGFPFARAV 289

90 EYIAD-CDRVLYLIDTVLVRDGLKP---LMD-----IDLGNWVGACIDLFVERQEGY 138

290 FYAAEICCGLEDLHRRIRIYRDLKPENILLDDHGIRISDLG-----LAVHVPGGQTI 342

139 KQKIG-----MADGEYRN-----AGVLLINL-----KKRRHDIFKMSC 173

343 KGVHGVGYMAPEVYANNEYTSPPDMALGCLLEYEMIAQSPPQORKKIKREEVERLYK 402

174 EWVEQYKDYMOYODDILNGLEK-----GSCVYANSRPFNPNVAFMANGF-- 220

403 EVAEETDFSSQARSLCSQLSKDPAERLGGRGGARREKHEPLFKINFRILGAGMLE 462

221 -----ASRHTDPLIDRTNTAM-----PVAVSHY-----CGSAK-PWHRDCTWGAER 262

463 PPKPDPQAIYCKDVIDIEQFSVKGVDLEPTDQDQKFAFGSVSIPQONEV--ETEC 520

263 FTFL--AGSLITYPE--EMRGIKLAVPPTKMLQR 292

521 FDELNVFGLDGVSVPDLDMKGQFTAPPKKGLLOR 554

Search completed: December 2, 2002, 11:56:27

Job time : 8.44739 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 26.2849 Seconds
(without alignments)
2398.732 Million cell updates/sec

Title: US-10-007-267-4

Perfect score: 1657

Sequence: 1 MDIVFAADDNYAAYLCVAAK.....KCMLORMKKLSARFLRKIV 306

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1657	100.0	306	2	Q50948	Q50948 neisseria g
2	1631.5	98.5	307	2	Q9REX5	Q9REX5 neisseria g
3	1551.5	93.6	311	2	P96945	P96945 neisseria m
4	1550.5	93.6	311	2	Q93EK7	Q93EK7 neisseria m
5	572	34.5	302	2	Q9L6B2	Q9L6B2 pasteurella
6	572	34.5	302	16	Q9CLSO	Q9CLSO pasteurella
7	284	17.1	57	2	Q9REX6	Q9REX6 neisseria g
8	278	16.8	60	2	Q9RGNO	Q9RGNO neisseria g
9	269.5	16.3	269	2	Q9L7A2	Q9L7A2 haemophilus
10	257	15.5	314	2	Q9AHB5	Q9AHB5 streptococc
11	255	15.4	630	2	Q48480	Q48480 klebsiella
12	243	14.7	314	2	Q8VU29	Q8VU29 streptococc
13	239	14.4	406	16	Q97P76	Q97P76 streptococc
14	233	14.1	398	16	Q97P77	Q97P77 streptococc
15	230.5	13.9	290	2	Q68207	Q68207 escherichia
16	220	13.3	335	2	Q9ZIT4	Q9ZIT4 escherichia

17	220	13.3	335	16	Q8XDC0	Q8XDC0 escherichia
18	219.5	13.2	273	2	Q93RH6	Q93RH6 streptococc
19	219.5	13.2	635	2	Q48484	Q48484 klebsiella
20	218.5	13.2	342	2	Q9ZIS1	Q9ZIS1 escherichia
21	217.5	13.1	338	2	Q9R9D1	Q9R9D1 escherichia
22	216.5	13.1	338	2	Q9ZIS5	Q9ZIS5 escherichia
23	212.5	12.8	341	2	Q9ZIS8	Q9ZIS8 escherichia
24	210	12.7	337	16	Q8Z2G3	Q8Z2G3 salmonella
25	201.5	12.2	404	16	Q97P73	Q97P73 streptococc
26	197.5	11.9	309	2	Q52832	Q52832 rhizobium 1
27	194	11.7	331	2	Q9R9D2	Q9R9D2 escherichia
28	193	11.6	331	2	Q9ZIS6	Q9ZIS6 escherichia
29	190.5	11.5	346	10	Q04253	Q04253 arbidopsis
30	190	11.5	517	16	Q9ZD68	Q9ZD68 rickettsia
31	189	11.4	339	2	Q68205	Q68205 escherichia
32	187	11.3	338	2	Q68204	Q68204 escherichia
33	186	11.2	361	10	Q9FWY9	Q9FWY9 arbidopsis
34	184	11.1	572	10	Q9LF35	Q9LF35 arbidopsis
35	183	11.0	510	10	Q9MAB8	Q9MAB8 arbidopsis
36	179.5	10.8	814	16	Q97P75	Q97P75 streptococc
37	177	10.7	329	2	Q9ZNA6	Q9ZNA6 erythrobact
38	177	10.7	337	16	Q8XDC2	Q8XDC2 escherichia
39	176.5	10.7	696	16	Q97P72	Q97P72 streptococc
40	171.5	10.4	357	10	Q9LZJ9	Q9LZJ9 arbidopsis
41	171.5	10.4	361	10	Q8VYF4	Q8VYF4 arbidopsis
42	171	10.3	682	2	Q9AEU2	Q9AEU2 streptococc
43	170	10.3	337	16	Q92V02	Q92V02 rhizobium m
44	169.5	10.2	351	10	Q9M8J2	Q9M8J2 arbidopsis
45	168	10.1	336	16	Q8Z2G2	Q8Z2G2 salmonella

ALIGNMENTS

RESULT 1	Q50948	PRELIMINARY:	PRT:	306 AA.
ID	Q50948			
AC	Q50948:			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Glycosyl transferase.			
DE	LGTC.			
OS	Neisseria gonorrhoeae.			
OC	Bacteria: Proteobacteria: beta subdivision; Neisseriaceae; Neisseria.			
OX	NCBI_TaxID=485:			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=P62:			
RX	MEDLINE=95053752; PubMed=7964493;			
RA	Gotschlich E.C.:			
RT	"Genetic locus for the biosynthesis of the variable portion of			
RT	Neisseria gonorrhoeae lipooligosaccharide.";			
RT	J. Exp. Med. 180:2181-2190(1994).			
DR	EMBL: U14554; AAA68011.1; "			
DR	InterPro: IPR002495; GT_8.			
DR	Pfam: PF01501; Glyco_transf_8; 1.			
KW	Transferase.			
SQ	SEQUENCE 306 AA: 34857 MW: 17855208FD88AACB CRC64:			
Query Match	100.0%; Score 1657; DB 2; Length 306;			
Best local Similarity	100.0%; Pred. No. 9.4e-142;			
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 MDIVFAADDNYAAYLCVAAKSVAAHPDIEIRHYVIDAGISEENRAVAANLRGGGINRF 60			
DB	1 MDIVFAADDNYAAYLCVAAKSVAAHPDIEIRHYVIDAGISEENRAVAANLRGGGINRF 60			
QY	61 IDVNPDPAGFPLNTHISTTYARLKLGXYIADCKVLYLDPDVLVROGLKLPMDTG 120			
DB	61 IDVNPDPAGFPLNTHISTTYARLKLGXYIADCKVLYLDPDVLVROGLKLPMDTG 120			
QY	121 GNVGACIDLEVEROBGYKOKIGMADGEYFFNAGVLLINLKKRRRHDIFKSCWEVQYK 180			

```

Db 121 GNMWVACIDLFEVERQGYKQKIGMADGEYFNAGVLLINKKWRHRDIFKMSCEWEQYK 180
      |||
QY 181 DVWQYODDOLNLGFKGVCYANSRNPMTNFAFANGASRHTDPLVDRTNTAMPVA 240
      |||
Db 181 DVWQYODDOLNLGFKGVCYANSRNPMTNFAFANGASRHTDPLVDRTNTAMPVA 240
      |||
QY 241 VSHYCGSAKPMHRDCTVWGAERFTELAGSLTTPPEEMRGKLAVPPTKCMLORRKKLSAR 300
      |||
Db 241 VSHYCGSAKPMHRDCTVWGAERFTELAGSLTTPPEEMRGKLAVPPTKCMLORRKKLSAR 300
      |||
QY 301 FLRKITY 306
      |||
Db 301 FLRKITY 306
      |||

RESULT 2
Q9REX5 PRELIMINARY; PRT; 307 AA.
ID Q9REX5;
AC Q9REX5;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE LgtC.
GN Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RA Balthazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT "Mutations in the lgt operon influence serum-resistance in gonococci.";
RT gonococci";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF208062; AAF20991.1; -.
DR InterPro; IPR002495; GT_8.
DR Pfam; PF01501; Glyco_transf_8; 1.
SQ SEQUENCE 307 AA; 34903 MW; 02FB7015CB1CE635 CRC64;

```

```

Query Match          98.5%; Score 1631.5; DB 2; Length 307;
Best Local Similarity 99.0%; Pred. No. 1.9e-139;
Matches 304; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

```

```

QY 1 MDIVFAADDNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAVAANLR-GGGNIR 59
      |||
Db 1 MDIVFAADDNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAVAANLRGGGGNIR 60
      |||
QY 60 FIDVNEDEFGFPLNIRHISITTYARLKIGEYIADCDKYLVDTPVLRDGLKPLMDTDL 119
      |||
Db 61 FIDVNEDEFGFPLNIRHISITTYARLKIGEYIADCDKYLVDTPVLRDGLKPLMDTDL 120
      |||
QY 120 GGNWVACIDLFEVERQGYKQKIGMADGEYFNAGVLLINKKWRHRDIFKMSCEWEQY 179
      |||
Db 121 GGNWVACIDLFEVERQGYKQKIGMADGEYFNAGVLLINKKWRHRDIFKMSCEWEQY 180
      |||
QY 180 KDVMOYODDOLNLGFKGVCYANSRNPMTNFAFANGASRHTDPLVDRTNTAMPV 239
      |||
Db 181 KDVMOYODDOLNLGFKGVCYANSRNPMTNFAFANGASRHTDPLVDRTNTAMPV 240
      |||
QY 240 AVSHYCGSAKPMHRDCTVWGAERFTELAGSLTTPPEEMRGKLAVPPTKCMLORRKKLSA 299
      |||
Db 241 AVSHYCGSAKPMHRDCTVWGAERFTELAGSLTTPPEEMRGKLAVPPTKCMLORRKKLSA 300
      |||
QY 300 FLRKITY 306
      |||
Db 301 FLRKITY 307
      |||

```

```

RESULT 3
P96945 PRELIMINARY; PRT; 311 AA.
ID P96945;
AC P96945;

```

```

DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Glycosyl transferase.
GN LgtC.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=126E;
RX MEDLINE=20055626; PubMed=10589709;
RA Jennings M.P., Srikanta Y.N., Moxon E.R., Kramer M., Poolman J.T.,
RA Kuipers B., van der Ley P.;
RT "The genetic basis of the phase variation repertoire of lipopolysaccharide immunotypes in Neisseria meningitidis.";
RT Microbiology 145:3013-3021(1999).
RL EMBL; U65788; AAB48385.1; -.
DR InterPro; IPR002495; GT_8.
DR Pfam; PF01501; Glyco_transf_8; 1.
DR Transferase.
KW SEQUENCE 311 AA; 35743 MW; A33D91283857750B CRC64;

```

```

Query Match          93.6%; Score 1551.5; DB 2; Length 311;
Best Local Similarity 93.9%; Pred. No. 3.4e-132;
Matches 292; Conservative 2; Mismatches 12; Indels 5; Gaps 2;

```

```

QY 1 MDIVFAADDNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAVAANLR-GGGNIR 59
      |||
Db 1 MDIVFAADDNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAVAANLRGGGGNIR 60
      |||
QY 60 FIDVNEDEFGFPLNIRHISITTYARLKIGEYIADCDKYLVDTPVLRDGLKPLMDTDL 119
      |||
Db 61 FIDVNEDEFGFPLNIRHISITTYARLKIGEYIADCDKYLVDTPVLRDGLKPLMDTDL 120
      |||
QY 120 GGNWVACIDLFEVERQGYKQKIGMADGEYFNAGVLLINKKWRHRDIFKMSCEWEQY 179
      |||
Db 121 GNMWVACIDLFEVERQGYKQKIGMADGEYFNAGVLLINKKWRHRDIFKMSCEWEQY 180
      |||
QY 180 KDVMOYODDOLNLGFKGVCYANSRNPMTNFAFANGASRHTDPLVDRTNTAMPV 239
      |||
Db 181 KDVMOYODDOLNLGFKGVCYANSRNPMTNFAFANGASRHTDPLVDRTNTAMPV 240
      |||
QY 240 AVSHYCGSAKPMHRDCTVWGAERFTELAGSLTTPPEEMRGKLAVPPTKCMLORRKKLSA 295
      |||
Db 241 AVSHYCGSAKPMHRDCTVWGAERFTELAGSLTTPPEEMRGKLAVPPTKCMLORRKKLSA 300
      |||
QY 296 KLSARFLRKITY 306
      |||
Db 301 KLSARFLRKITY 311
      |||

```

```

RESULT 4
Q93EK7 PRELIMINARY; PRT; 311 AA.
ID Q93EK7;
AC Q93EK7;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE LgtC.
GN Neisseria meningitidis.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M978;
RX MEDLINE=21467954; PubMed=11583844;
RA Zhu P., Kutuch M.J., Tsai C.-M.;
RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";
RT FEMS Microbiol. Lett. 203:173-177(2001).
RL EMBL; AF355193; AAL12839.1; -.

```

DR InterPro: IPR002495; GT_8.
 Pfam: PF01501; Glyco_transf_8; 1.
 SQ SEQUENCE 311 AA; 35738 MW; A6EDE3F85AC5B5D CRC64;

Query Match 93.6%; Score 1550.5; DB 2; Length 311;
 Best Local Similarity 93.9%; Pred. No. 4.2e-132;
 Matches 292; Conservative 2; Mismatches 12; Indels 5; Gaps 2;

QY 1 MDIVFADADNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVALNR-GGNGIR 59
 1 MDIVFADADNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVALNRGGGNGIR 60
 DB 60 FIDVNEDEPAGPPLNIRHISITTYARLKGEYIADCDKVLVLDVLRDGLKPLMDTD 119
 61 FIDVNEDEPAGPPLNIRHISITTYARLKGEYIADCDKVLVLDVLRDGLKPLMDTD 120
 QY 120 GGNWVGACIDLFVEROEGYKOKIGMADGEYFNAGVLLINLKKWRHDIFFKMSCVEVEQ 179
 121 GGNWVGACIDLFVEROEGYKOKIGMADGEYFNAGVLLINLKKWRHDIFFKMSCVEVEQ 180
 QY 180 KQVMQYQDODILNGLFGKGCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMPY 239
 181 KQVMQYQDODILNGLFGKGCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMPY 240
 DB 240 AVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEWKGKLAVP---PTKCMLORWK 295
 241 AVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEWKGKLAVP---PTKCMLORWK 300
 QY 296 KLSARFLRKTY 306
 301 KLSARFLRKTY 311

RESULT 5
 Q9L6B2 PRELIMINARY; PRT: 302 AA.

AC Q9L6B2; PRELIMINARY; PRT: 302 AA.
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Putative glycosyl transferase.
 GN LGTC.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_Taxid=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fuller T.E., Kennedy M.J., Lowery D.E.;
 RT Identification of Pasteurella multocida virulence genes in a
 RT septicemic mouse model using signature-tagged mutagenesis";
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF237927; AAF68413.1; -
 DR InterPro: IPR002495; GT_8.
 DR Pfam: PF01501; Glyco_transf_8; 1.
 DR Transferase.
 SQ SEQUENCE 302 AA; 35688 MW; C987B3F77FB158 CRC64;

Query Match 34.5%; Score 572; DB 2; Length 302;

Best Local Similarity 38.9%; Pred. No. 1.4e-43;

Matches 121; Conservative 61; Mismatches 115; Indels 14; Gaps 7;

QY 1 MDIVFADADNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVALNR-GGNGIR 59
 1 MDIVFADADNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVALNRGGGNGIR 59
 DB 1 MDIVFADADNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVALNRGGGNGIR 59
 QY 60 FIDVNEDEPAGPPLNIRHISITTYARLKGEYIAD-CDKVLVLDVLRDGLKPLMDTD 118
 61 FIDVNEDEPAGPPLNIRHISITTYARLKGEYIAD-CDKVLVLDVLRDGLKPLMDTD 119
 DB 60 FIDVNEDEPAGPPLNIRHISITTYARLKGEYIAD-CDKVLVLDVLRDGLKPLMDTD 118
 61 FIDVNEDEPAGPPLNIRHISITTYARLKGEYIAD-CDKVLVLDVLRDGLKPLMDTD 119
 QY 119 LGGNWWGACIDLFVEROEG-YKOKIGMADGEYFNAGVLLINLKKWRHDIFFKMSCVEVE 177
 120 LGGNWWGACIDLFVEROEG-YKOKIGMADGEYFNAGVLLINLKKWRHDIFFKMSCVEVE 179
 DB 120 LGGNWWGACIDLFVEROEG-YKOKIGMADGEYFNAGVLLINLKKWRHDIFFKMSCVEVE 179

QY 178 QYKVMQYQDODILNGLFGKGCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMP 237
 179 QYKVMQYQDODILNGLFGKGCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMP 237
 DB 180 MYPNOMIYQDODILNGLFGKGCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMP 237
 181 MYPNOMIYQDODILNGLFGKGCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMP 237
 QY 238 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEWKGKLAVP---PTKCMLORWK 295
 239 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEWKGKLAVP---PTKCMLORWK 295
 DB 238 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEWKGKLAVP---PTKCMLORWK 295
 239 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEWKGKLAVP---PTKCMLORWK 295

QY 296 KLSARFLRKTY 306
 297 KLSARFLRKTY 302

RESULT 6

Q9CL50 PRELIMINARY; PRT: 302 AA.

AC Q9CL50; PRELIMINARY; PRT: 302 AA.
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Hypothetical protein PM139.
 GN PM139.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_Taxid=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AE006155; AKK03223.1; -
 DR InterPro: IPR002495; GT_8.
 DR Pfam: PF01501; Glyco_transf_8; 1.
 DR Hypothetical protein: Complete proteome.
 SQ SEQUENCE 302 AA; 35716 MW; 6C7264ED8F033C00 CRC64;

Query Match 34.5%; Score 572; DB 16; Length 302;

Best Local Similarity 38.9%; Pred. No. 1.4e-43;

Matches 121; Conservative 61; Mismatches 115; Indels 14; Gaps 7;

QY 1 MDIVFADADNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVALNR-GGNGIR 59
 1 MDIVFADADNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVALNRGGGNGIR 59
 DB 1 MDIVFADADNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVALNRGGGNGIR 59
 QY 60 FIDVNEDEPAGPPLNIRHISITTYARLKGEYIAD-CDKVLVLDVLRDGLKPLMDTD 118
 61 FIDVNEDEPAGPPLNIRHISITTYARLKGEYIAD-CDKVLVLDVLRDGLKPLMDTD 119
 DB 60 FIDVNEDEPAGPPLNIRHISITTYARLKGEYIAD-CDKVLVLDVLRDGLKPLMDTD 118
 61 FIDVNEDEPAGPPLNIRHISITTYARLKGEYIAD-CDKVLVLDVLRDGLKPLMDTD 119
 QY 119 LGGNWWGACIDLFVEROEG-YKOKIGMADGEYFNAGVLLINLKKWRHDIFFKMSCVEVE 177
 120 LGGNWWGACIDLFVEROEG-YKOKIGMADGEYFNAGVLLINLKKWRHDIFFKMSCVEVE 179
 DB 120 LGGNWWGACIDLFVEROEG-YKOKIGMADGEYFNAGVLLINLKKWRHDIFFKMSCVEVE 179
 QY 178 QYKVMQYQDODILNGLFGKGCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMP 237
 179 QYKVMQYQDODILNGLFGKGCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMP 237
 DB 180 MYPNOMIYQDODILNGLFGKGCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMP 237
 181 MYPNOMIYQDODILNGLFGKGCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMP 237
 QY 238 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEWKGKLAVP---PTKCMLORWK 295
 239 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEWKGKLAVP---PTKCMLORWK 295
 DB 238 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEWKGKLAVP---PTKCMLORWK 295
 239 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEWKGKLAVP---PTKCMLORWK 295
 QY 296 KLSARFLRKTY 306
 297 KLSARFLRKTY 302
 DB 292 RIRYKFRYQY 302

RESULT 7

Q9REX6 PRELIMINARY; PRT: 57 AA.

ID Q9REX6
 Q9REX6;

DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE LgtC.
CN LgtC.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA19;
RA Baltazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT "Mutations in the lgt operon influence serum-resistance in gonococci."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF208061; AAF20990.1; -
DR InterPro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
SQ SEQUENCE 57 AA; 5934 MW; 392247E1A7C1135D CRC64;

Query Match 17.1%; Score 284; DB 2; Length 57;
Best Local Similarity 100.0%; Pred. No. 1.6e-18;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVFADDNYAAYLCVAAKSVEAAHPDTEIRFHVLDAGISEENRAVAANLRGG 56
Db 1 MDIVFADDNYAAYLCVAAKSVEAAHPDTEIRFHVLDAGISEENRAVAANLRGG 56

RESULT 8
Q9RGNO PRELIMINARY; PRT; 60 AA.

AC Q9RGNO;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE Glycosyltransferase.
GN LgtC.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1291;
RX MEDLINE=20305049; PubMed=10844691;
RA Harvey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
RT Phillips N.J., Apicella M.A., Balke M.S.;
RT "Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein receptor on human sperm."
RL MOL. Microbiol. 36:1059-1070(2000).
DR EMBL: AF121135; AAF14361.1; -
DR InterPro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
KM transferase.
SQ SEQUENCE 60 AA; 6304 MW; D28468118F7E1A7 CRC64;

Query Match 16.8%; Score 278; DB 2; Length 60;
Best Local Similarity 100.0%; Pred. No. 5.9e-18;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVFADDNYAAYLCVAAKSVEAAHPDTEIRFHVLDAGISEENRAVAANLRGG 55
Db 1 MDIVFADDNYAAYLCVAAKSVEAAHPDTEIRFHVLDAGISEENRAVAANLRGG 55

RESULT 9

Q9L7A2 PRELIMINARY; PRT; 269 AA.

AC Q9L7A2;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
DE Putative glycosyl transferase.

OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang J., Hanson E., Munson R.S. Jr.;
RT "Putative glycosyl transferase."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF224467; AAF28363.1; -
DR InterPro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
KM transferase.
SQ SEQUENCE 269 AA; 32334 MW; 1F2E25A243AAC64 CRC64;

Query Match 16.3%; Score 269.5; DB 2; Length 269;
Best Local Similarity 30.8%; Pred. No. 2.9e-16;
Matches 78; Conservative 43; Mismatches 111; Indels 21; Gaps 7;

QY 1 MDIVFADDNYAAYLCVAAKSVEAAHPDTEIRFHVLDAGISEENRAVAANLRG-GGNIR 59
Db 8 MNIYLAANOSYSEYILTTKISYTL--HNKHIREYLLNRDYPTEMDILNKKLKLNSSEIT 65
QY 60 FIDVNPEDFAFPPLNIRHISITTYARIKGEYIADCXVLYLDPVLYRDLKPLMDTDL 119
Db 66 DIKVTNDTIKFNFTYSHISSDTEFFRYISDFI-EDQKVIYLDADIVNGSLFELYQTDI 124
QY 120 GGNWVGACIDLFVEROGEYQKIGMADGEYFNAGVLLILKKWRHDFKMSCENEYQ 179
Db 125 SNYFLAAVKDISEK-----IYNNHIFNAGMLLNKKWRHNTIQFCLISEKY 175
QY 180 KDVMQYODDILNGLFKGGVYCANSRFNM-PTNYAFMANGFASRRDPLDRTYAMP 238
Db 176 INSLPADQSIILNIRKDKLKNRCYINLYIGDYLFFKYG-----KTYLEDLGETIP 229
QY 239 VAVSHYCGSAPW 251
Db 230 LIH-HYNTAKPW 241

RESULT 10

Q9AHB5 PRELIMINARY; PRT; 314 AA.

AC Q9AHB5;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE wcin (CpsP) (Putative galactosyl transferase) (Putative galactosyl transferase Cps6ap).
GN wcin OR CpsP OR Cps6ap.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21116929; PubMed=11179285;
RA Jiang S.M., Wang L., Reeves P.R.;
RT "Molecular characterization of Streptococcus pneumoniae type 4, 6B, 8, and 18C capsular polysaccharide gene clusters."
RL Infect. Immun. 69:1244-1255(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PN93/1413;
RA Griffiths D.B., Hall L.M.C.;
RT "The Cps locus of Streptococcus pneumoniae serotype 6B, genes between cpsB and cpsL."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

RP [3]
RC SEQUENCE FROM N.A.
RC STRAIN=SSISP 64/L.
RA Griffiths D.B., Hall L.M.C.;
RT "The Capsular Polysaccharide Locus of Streptococcus pneumoniae

```

RT Serotype 6A."
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-PN93/Y008:
RA Griffiths D.B., Hall L.M.C.;
RT "The capsular polysaccharide locus of Streptococcus pneumoniae
RT serotype 6A."
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF316640; AAK20686.1; -
DR EMBL: AF298581; AAL6843.1; -
DR EMBL: AF246898; AAL6842.1; -
DR EMBL: AF078347; AAL82783.1; -
DR Interpro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
DR Transferase.
SQ SEQUENCE 314 AA: 36635 MW: 2073B87BF3D317C5 CRC64:

Query Match 15.5%; Score 257; DB 2; Length 314;
Best Local Similarity 25.0%; Pred. No. 4,8e-15;
Matches 83; Conservative 66; Mismatches 125; Indels 58; Gaps 13;

QY 1 MDIVFADNDNYAAYLCVAKSYEAHNPDEIRFHVLDAGISEENRAAV--AANLRGGGNI 58
DB 1 MDIVATDNNFVDVLSASIKSLYTTNSDLNMLIADKVSDBRKNKIRLSQFAQREI 60
QY 59 RFDVNPEDFAGFPLNIRIISITTYARLKGLEYI-ADCKVLYLDVDLVROGLKPLMDT 117
DB 61 NMIE-NVE--IFPKLHLDGSISSFSRFLGSLVPSMSKVLXLDSDIIVMDSLRKIFDI 117
QY 118 DLGMMVGCACIDLFVEROGYKOKIGMADGEYFVNAVLLINLKKRRHDFRMSCEWE 177
DB 118 DEKGKILYGVNDTF--NKEYKQVIGIPIDKPMFNAGVMLINLEARNNVBERFLQVIO 174
QY 178 QYKDWQVQYODDILNGLFGGVCYANSRNFMPNTNFAFANGFASHHTPLYLDRT---- 233
DB 175 KFNGLTLOGDLVCLN-----AVLY--NSFVLPPEYNYMT-----IFEDLYEEM 217
QY 234 -----NTAMPVAVSHYCG---SAKPMHDCYVWGAERFTE-LAGSLTTY 273
DB 218 IYFKRPINYSKEEIKNAERIYLRHFTTSLSKRPWQSGSNVAHIDQIKKYEYSKRV 277
QY 274 PEEMRGKLAVP-PTKCM-----QRMRRKL 297
DB 278 KESILKIVQIKPKKCSYVFLGIQSKFRPKL 309

RESULT 11
Q48480 PRELIMINARY; PRT; 630 AA.
AC Q48480;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE WbM protein.
GN WbM.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OC NCBL_TaxID=573;
OX NCBL [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE O8;
RX MEDLINE=96359375; PubMed=8752339;
RA "Clonally diverse rfb gene clusters are involved in expression of a
RT family of related D-galactan O antigens in Klebsiella species."
RT J. Bacteriol. 178:5205-5214(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE O8;
RX MEDLINE=97158136; PubMed=9004408;
RA Reeves P.R., Hobbs M., Valvano M.A., Skurnik M., Whitfield C.,

```

```

RA Coplin D., Kido N., Klena J., Maskell D., Rietz C.R., Rick P.D.;
RT "Bacterial polysaccharide synthesis and gene nomenclature."
RT Trends Microbiol. 4:495-503(1996).
DR EMBL: L41518; AAC98407.1; -
DR Interpro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
SQ SEQUENCE 630 AA: 72599 MW: BA8DF92A6D1193AA CRC64:

Query Match 15.4%; Score 255; DB 2; Length 630;
Best Local Similarity 27.6%; Pred. No. 1.9e-14;
Matches 79; Conservative 47; Mismatches 108; Indels 52; Gaps 10;

QY 3 IYFADNDNYAAYLCVAKSYEAHNPDEIRFH--VLDAISEENRAVANLRGGGNI-- 58
DB 279 IYISFDNNYAVSGALINSI-VRHADKNKNYDVLNVNLSNKKRLHLVAQGSINSL 337
QY 59 RFDVNPEDFAGFPLNIRIISITTYARLKGLEYIADCDKVLVDVLYROGLKPLMDT 118
DB 338 RFDVNPFTETISAVHTRAFHSASTYARLFPOLFRSYDKVVFIDSDPTVVKADLATLMDVD 397
QY 119 LCGMNVGACIDLFVE-----ROEGYKOKIGMADGEYFVNAVLLINL 161
DB 398 IGTNLVAAYKIDYMEGFVFGAMSESDGVMPAKETLQTLGTNTDPEYFOAGITVENVG 457
QY 162 KMRHDFRMSCEWEQYKDVQYODDILNGLFGGVCYANSRNFMPNTN-APFANGF 220
DB 458 QWVKEDTFSLMATLAKK--YWFLLDQDILNNKVFEG-----RVKFLLENNVYHGNG- 507
QY 221 ASRHNDPLYL-----RINTAMPVAVSHYCGSAKKPMHRD 254
DB 508 ---NTDFFPNLKFSTPMRFLQARSNPKM---IHYAGENKPMNTD 546

RESULT 12
Q8VU29 PRELIMINARY; PRT; 314 AA.
AC Q8VU29;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Putative galactosyl transferase.
GN CPSP.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OC NCBL_TaxID=1313;
OX NCBL [1]
RP SEQUENCE FROM N.A.
RC STRAIN-361;
RA Griffiths D.B., Hall L.M.C.;
RT "The Capsular Polysaccharide Locus of Streptococcus pneumoniae
RT Serotype 6B."
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF246897; AAL68409.1; -
DR Interpro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
DR Transferase.
SQ SEQUENCE 314 AA: 36782 MW: FBB2E997F3E1591 CRC64:

Query Match 14.7%; Score 243; DB 2; Length 314;
Best Local Similarity 24.1%; Pred. No. 8.9e-14;
Matches 80; Conservative 63; Mismatches 117; Indels 72; Gaps 12;

QY 1 MDIVFADNDNYAAYLCVAKSYEAHNPDEIRFHVLDAGISEENRAAV--AANLRGGGNI 58
DB 1 MDIVATDNNFVDVLSASIKSLYTTNSDLNMLIADKVSDBRKNKIRLSQFAQREI 60
QY 59 RFDVNPEDFAGFPLNIRIISITTYARLKGLEYI-ADCKVLYLDVDLVROGLKPLMDT 117
DB 61 NMIE-NVE--IFPKLHLDGSISSFSRFLGSLVPSMSKVLXLDSDIIVMDSLRKIFDI 117
QY 118 DLGMMVGCACIDLFVEROGYKOKIGMADGEYFVNAVLLINLKKRRHDFRMSCEWE 177

```


OC	Escherichia.
XX	NCBI_TaxID=562;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=F632;
RX	MEDLINE=98204873; PubMed=9535865;
RA	Heinrichs D.E., Monteiro M.A., Perry M.B., Whitfield C.;
RT	"The assembly system for the lipopolysaccharide R2 core-type of
Rt	Escherichia coli is a hybrid of those found in Escherichia coli K-12
Rt	and Salmonella enterica. Structure and function of the R2 waak and
Rt	waal homologs." ;
RL	J. Biol. Chem. 273:8849-8859(1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=F632;
RX	MEDLINE=99009352; PubMed=9791168;
RA	Heinrichs D.E., Yelton J.A., Whitfield C.;
RT	"Molecular basis for structural diversity in the core regions of the
Rt	lipopolysaccharides of escherichia coli and salmonella enterica."
RL	Mol. Microbiol. 30:221-232(1998).
DR	EMBL: AF019375; AAC69655.1; -
DR	InterPro: IPR002495; GT_8.
SQ	Pfam: PF01501; Glyco_transf_8; 1.
	Seqence 290 AA: 33801 MW; 3A1A6FAF40C7AF7B CRC64;
Query Match	13.9%; Score 230.5; DB 2: Length 290:
Best Local Similarity	28.8%; Pred. No.1.le-12;
Matches	74; Conservative 46; Mismatches 100; Indels 37; Gaps 11.
OY	17 VAKSVEAHAPDTEIRFHVLADIGISEENRAVAANLRGG--NTRFI---DVNPDEFA 69
DB	11 : : : : : : : : : : : : : : :
DB	5 VALYSITSNNRHINLHPIHLISGI-EKCKSAFYEELEGPNTSISVYYIIDRKDINDPDLV 63
OY	70 -GFLAIRIHISTTAYARLKLGXYIAD-CDKYLYITDVLVRGCLRPLMTDTLGGMWGCAC 127
DB	64 LGIPV-----STGLRFLEVIDDKINKLLIYLDCDYHCNCPLELDVDYNLNAD-IACV 115
OY	128 IDLFERROEGYKOKIGMADGEYYFNAGVLLIMLKKWRHDIFKMSEWEQYKDWQOYOD 187
DB	116 IPDSFMGERVKAKKLDYGIEFIYTFNAGVAFINTSERKKNNITQKALEMINSGR-VYRYAD 174
OY	188 ODILNGLEFGGVGYVANSRENMPTVMFANGFASRRHTDPLYLDRTNAMP-VAVSHYCG 246
DB	175 ODVLIILLNGRVHYIYDKKYN-----NKTTLVSVCDEEQKMLPTIIIMHYVT 220
OY	247 SAKPMWNRDCTWGGAERF 263
DB	221 ONKPWYK---IFRAQNF 234

THIS PAGE BLANK (USPTO)

XX Claim 11: Fig 2e; 81pp; English.
 PS
 CC 5 Glycosyltransferases (AAR91311-15) are products of the lgt locus
 CC (AAR14061) of *Neisseria gonorrhoeae* strain F62. Glycosyltransferase
 CC LgtD (AAR91314) can be obtd. by expression of the lgtD coding
 CC sequence in recombinant host cells. A method for adding GalNAc
 CC or GlcNAc beta1-3 to Gal comprises contacting a reaction mixture
 CC congt. activated GalNAc or GlcNAc to an acceptor moiety comprising a
 CC Gal residue in the presence of LgtD. Oligosaccharides can be produced
 CC that, when attached to non-toxic lipids, are useful for *Neisseria*
 CC vaccine prepn. Blood group core oligosaccharides, and mimics of
 CC lacto-N-neotetraose, gangliosides and saccharide portions of
 CC globoglycolipids can also be produced using the enzymes.
 CC
 SQ Sequence 337 AA;

Query Match 100.0%; Score 1764; DB 17; Length 337;
 Best Local Similarity 100.0%; Pred. No. 3.6e-178;
 Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MDPVSVLICAYNAEKYFPAQSLAAVVGQTWNRNDILIVDGSSTDGTPATARRHQEDGR 60
 DB 1 MDPVSVLICAYNAEKYFPAQSLAAVVGQTWNRNDILIVDGSSTDGTPATARRHQEDGR 60
 OY 61 RIISNPRNLGFIASLNLIGDELAKSGGGEYIARTDADDIASPGWIKIYGEKEDRSIIA 120
 DB 61 RIISNPRNLGFIASLNLIGDELAKSGGGEYIARTDADDIASPGWIKIYGEKEDRSIIA 120
 OY 121 MGAMLEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIVAAPFGPNPIHNTMTIMRSVIDG 180
 DB 121 MGAMLEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIVAAPFGPNPIHNTMTIMRSVIDG 180
 OY 181 GLRFDPAIYHAEDYKFWYAGKIGRLAYYPEALVKYRFHQDOTSSKYNLQOORTAWKIKE 240
 DB 181 GLRFDPAIYHAEDYKFWYAGKIGRLAYYPEALVKYRFHQDOTSSKYNLQOORTAWKIKE 240
 OY 241 EIRAGWKAAGIAGVADCLNTYGLKSTAYALYEKALSODIGCLRFLVEYFLSEKYSYL 300
 DB 241 EIRAGWKAAGIAGVADCLNTYGLKSTAYALYEKALSODIGCLRFLVEYFLSEKYSYL 300
 OY 301 TDLDLFLTRVVRMKLFAAPQYRKILKKMLRPMKYRSY 337
 DB 301 TDLDLFLTRVVRMKLFAAPQYRKILKKMLRPMKYRSY 337

RESULT 2

AAW06579
 ID AAW06579 standard; Protein; 337 AA.

AC AAW06579;

DT 21-MAR-1997 (first entry)

DE Lipo-oligosaccharide gene-encoded protein.

XX Polyglycosyltransferase; N-acetylglucosaminyl transferase;
 KM N-acetylglucosaminyl transferase; lipo-oligosaccharide.
 XX
 OS *Neisseria gonorrhoeae* ATCC 33084.

XX WO9640971-A1.

XX 19-DEC-1996.

XX 03-JUN-1996; 96WO-US08323.

XX 07-JUN-1995; 95US-0478140.

XX (NEOS-) NEOS TECHNOLOGIES INC.

XX Buczala SL, Johnson KF, Roth S;

XX

DR WPI: 1997-052351/05.
 DR N-PSDB: AAT9230.
 XX
 PT Transfer of at least 2 saccharide units using
 PT polyglycosyltransferase - isolated from *N. gonorrhoeae*, catalyses
 PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
 PT single galactose moiety
 PS
 CC Disclosure; Fig 2E-F; 38pp; English.

XX A lipo-oligosaccharide-encoding gene region (AAT9230) of *Neisseria*
 CC *gonorrhoeae* ATCC 33084 includes coding sequences for 5 proteins
 CC (AAW06576-80), one of which (AAW06576) is a polyglycosyltransferase
 CC that catalyses the addition of GlcNAc and GalNAc disaccharides to
 CC a galactose moiety. The function of the other 4 proteins is not
 CC stated in the specification.

SQ Sequence 337 AA;

Query Match 99.8%; Score 1761; DB 18; Length 337;
 Best Local Similarity 99.7%; Pred. No. 7.5e-178;
 Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MDPVSVLICAYNAEKYFPAQSLAAVVGQTWNRNDILIVDGSSTDGTPATARRHQEDGR 60
 DB 1 MDPVSVLICAYNAEKYFPAQSLAAVVGQTWNRNDILIVDGSSTDGTPATARRHQEDGR 60
 OY 61 RIISNPRNLGFIASLNLIGDELAKSGGGEYIARTDADDIASPGWIKIYGEKEDRSIIA 120
 DB 61 RIISNPRNLGFIASLNLIGDELAKSGGGEYIARTDADDIASPGWIKIYGEKEDRSIIA 120
 OY 121 MGAMLEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIVAAPFGPNPIHNTMTIMRSVIDG 180
 DB 121 MGAMLEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIVAAPFGPNPIHNTMTIMRSVIDG 180
 OY 181 GLRFDPAIYHAEDYKFWYAGKIGRLAYYPEALVKYRFHQDOTSSKYNLQOORTAWKIKE 240
 DB 181 GLRFDPAIYHAEDYKFWYAGKIGRLAYYPEALVKYRFHQDOTSSKYNLQOORTAWKIKE 240
 OY 241 EIRAGWKAAGIAGVADCLNTYGLKSTAYALYEKALSODIGCLRFLVEYFLSEKYSYL 300
 DB 241 EIRAGWKAAGIAGVADCLNTYGLKSTAYALYEKALSODIGCLRFLVEYFLSEKYSYL 300
 OY 301 TDLDLFLTRVVRMKLFAAPQYRKILKKMLRPMKYRSY 337
 DB 301 TDLDLFLTRVVRMKLFAAPQYRKILKKMLRPMKYRSY 337

RESULT 3

AAR91311
 ID AAR91311 standard; Protein; 348 AA.

AC AAR91311;

DT 09-JUL-1996 (first entry)

DE *N. gonorrhoeae* glycosyltransferase LgtA.

XX Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
 KM vaccine.
 XX
 OS *Neisseria gonorrhoeae* strain F62.

XX WO9610086-A1.

XX 04-APR-1996.

XX 25-SEP-1995; 95WO-US12317.

XX 26-SEP-1994; 94US-0312387.

XX (UYRO) UNIV ROCKEFELLER.

XX

PI Gotschlich EC;
 XX
 DR WPI: 1996-200924/20.
 XX N-PSDB; AAT14061.
 XX
 PT Nucleic acids encoding glycosyl transferase(s) - used in the
 PT diagnosis of infection with Neisseria and for the biosynthesis of
 PT oligo:saccharide(s)
 XX
 PS Claim 8; Fig 2b; 81pp; English.
 XX
 CC 5 glycosyltransferases (AAR91311-15) are products of the 1gt locus
 CC (AAT14061) of Neisseria gonorrhoeae strain F62. Glycosyltransferase
 CC LgtA (AAR91311) can be obtained by expression of the lgtA coding
 CC sequence in recombinant host cells. A method for adding GalNAc
 CC or GlcNAc beta1-3 to Gal comprises contacting a reaction mixture
 CC conng. activated GalNAc or GlcNAc to an acceptor molecule comprising
 CC a Gal residue in the presence of LgtA. Oligosaccharides can be
 CC produced that, when attached to non-toxic lipids, are useful for
 CC Neisseria vaccine prepns. Blood group core oligosaccharides, and
 CC mimics of lacto-N-neotetraose, gangliosides and saccharide portions
 CC of globoglycolipids can also be produced using the enzymes.
 XX
 SQ Sequence 348 AA:
 Query Match 59.5%; Score 1049; DB 17; Length 348;
 Best Local Similarity 64.5%; Pred. No. 2.5e-102;
 Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;
 OY 1 MOPLVSVLICAYNAEKYFAQSLAAVVGOTWRNDLITVDGSGTGPATARHFOEDGRI 60
 DB 1 MOPLVSVLICAYNAEKYFAQSLAAVVGOTWRNDLITVDGSGTGPATARHFOEDGRI 60
 OY 61 RIISNPRNGFASINIGDELAKS--GGGEYIARTDADDIASPGITEKIVGEMEKDRSI 118
 DB 61 RIISNPRNGFASINIGDELAKS--GGGEYIARTDADDIASPGITEKIVGEMEKDRSI 120
 OY 119 IMGAWLEVLSEENKNSVAAIARNGAIWDKPTRHEDIYAAPFGNPJHNNMIMRSYI 178
 DB 121 IMGAWLEVLSEENKNSVAAIARNGAIWDKPTRHEDIYAAPFGNPJHNNMIMRSYI 180
 OY 179 DGLRDPAYIAHEDYKFEYEAAGLGRLAYYPEALVKYRFHODQTSKYNLQOARTWAKI 238
 DB 181 DGLRDPAYIAHEDYKFEYEAAGLGRLAYYPEALVKYRFHODQTSKYNLQOARTWAKI 240
 OY 239 KEETIRAGYKKAAGIAGACLVWGLKSTAYALYKALSGOIGCLRFLYEYFSLSEKY 298
 DB 241 KEETIRAGYKKAAGIAGACLVWGLKSTAYALYKALSGOIGCLRFLYEYFSLSEKY 300
 OY 299 SLTDLDFLTDRVMRKLFAPAQRKILKKMLR 330
 DB 301 PSGAMLDFAADGMRRLFTLRQYFGILYRLIK 332
 RESULT 4
 AAM06576
 ID AAM06576 standard; Protein; 348 AA.
 XX AAM06576;
 AC
 DT 21-MAR-1997 (first entry)
 XX
 DE Neisseria polyglucosyltransferase.
 XX
 KM Polyglucosyltransferase; N-acetylglucosaminyl transferase;
 KM N-acetylglucosaminyl transferase; lipo-oligosaccharide.
 XX
 OS Neisseria gonorrhoeae ATCC 33084.
 XX
 PN W09640971-A1.
 XX
 PD 19-DEC-1996.
 XX

PF 03-JUN-1996; 96WO-US08323.
 XX
 PR 07-JUN-1995; 95US-0478140.
 XX
 PA (NEOS-) NEOSE TECHNOLOGIES INC.
 XX
 PI Buczala SL, Johnson KF, Roth S;
 XX
 DR WPI: 1997-052351/05.
 DR N-PSDB; AAT49230.
 XX
 PT Transfer of at least 2 saccharide units using
 PT poly:glycosyltransferase - isolated from N. gonorrhoeae, catalyses
 PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
 PT single galactose moiety
 XX
 PS Disclosure; Fig 2A-C; 38pp; English.
 XX
 CC A novel polyglucosyltransferase (PGTase) (AAM06576) from Neisseria
 CC gonorrhoeae ATCC 33084 catalyses the stereospecific conjugation of
 CC 2 specific activated saccharide units (e.g. UDP-GlcNAc, UDP-GalNAc,
 CC UDP-Gal) to specific acceptors having a galactose moiety at a
 CC non-reducing end. It is the first PGTase reported to be capable of
 CC transfer of more than one different saccharide moiety. The PGTase
 CC is encoded by nucleotides 445-1488 of a lipo-oligosaccharide gene
 CC (AAT49230). It can be produced in transformed host cells and used in
 CC oligosaccharide prodn.
 XX
 SQ Sequence 348 AA:
 Query Match 59.3%; Score 1046; DB 18; Length 348;
 Best Local Similarity 64.2%; Pred. No. 5.3e-102;
 Matches 213; Conservative 34; Mismatches 83; Indels 2; Gaps 1;
 OY 1 MOPLVSVLICAYNAEKYFAQSLAAVVGOTWRNDLITVDGSGTGPATARHFOEDGRI 60
 DB 1 MOPLVSVLICAYNAEKYFAQSLAAVVGOTWRNDLITVDGSGTGPATARHFOEDGRI 60
 OY 61 RIISNPRNGFASINIGDELAKS--GGGEYIARTDADDIASPGITEKIVGEMEKDRSI 118
 DB 61 RIISNPRNGFASINIGDELAKS--GGGEYIARTDADDIASPGITEKIVGEMEKDRSI 120
 OY 119 IMGAWLEVLSEENKNSVAAIARNGAIWDKPTRHEDIYAAPFGNPJHNNMIMRSYI 178
 DB 121 IMGAWLEVLSEENKNSVAAIARNGAIWDKPTRHEDIYAAPFGNPJHNNMIMRSYI 180
 OY 179 DGLRDPAYIAHEDYKFEYEAAGLGRLAYYPEALVKYRFHODQTSKYNLQOARTWAKI 238
 DB 181 DGLRDPAYIAHEDYKFEYEAAGLGRLAYYPEALVKYRFHODQTSKYNLQOARTWAKI 240
 OY 239 KEETIRAGYKKAAGIAGACLVWGLKSTAYALYKALSGOIGCLRFLYEYFSLSEKY 298
 DB 241 KEETIRAGYKKAAGIAGACLVWGLKSTAYALYKALSGOIGCLRFLYEYFSLSEKY 300
 OY 299 SLTDLDFLTDRVMRKLFAPAQRKILKKMLR 330
 DB 301 PSGAMLDFAADGMRRLFTLRQYFGILYRLIK 332
 RESULT 5
 AAU072923
 ID AAU072923 standard; Protein; 346 AA.
 XX AAU072923;
 AC
 DT 12-MAR-2002 (first entry)
 XX
 DE Neisseria meningitidis virulence protein #13.
 XX
 KM Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
 KM infection; Gram-negative bacteria; antimicrobial.
 XX
 OS Neisseria meningitidis.
 OS

```
XX PN WO200185772-A2.
XX PD 15-NOV-2001.
XX PF 08-MAY-2001; 2001WO-GB02003.
XX PR 08-MAY-2000; 2000GB-0011108.
XX PA (MICR-) MICROSCIENCE LTD.
XX PI Tang C;
XX DR WPI; 2002-066593/09.
XX DR N-PSDB; AAS97208.
XX PT New peptide encoded by operon including virulence genes of Neisseria
XX PT meningitidis, useful as vaccine component for treating or preventing
XX PT meningitis and for identifying antimicrobial drug
XX PS Claim 4; Page 60-61; 423pp; English.
XX CC The invention relates to a peptide (I) encoded by an operon (II) of
XX CC Neisseria meningitidis including virulence genes, or a related molecule
XX CC having a 40% sequence similarity at the peptide or nucleotide level in a
XX CC Gram-negative bacterium, or its functional fragment, for therapeutic or
XX CC diagnostic use. (I) and (II) are useful in the manufacture of a
XX CC medicament for treating or preventing a condition (e.g., meningitis)
XX CC associated with infection by Neisseria or Gram-negative bacteria. The
XX CC product is useful for veterinary treatment and in a screening assay for
XX CC the identification of an antimicrobial drug. The vaccines have
XX CC prophylactic applications. AAU72911-AAU73014 represent N. meningitidis
XX CC virulence proteins of the invention.
XX SQ Sequence 346 AA:
XX
XX Query Match 58.7%; Score 1036; DB 23; Length 346;
XX Best Local Similarity 63.0%; Pred. No. 6e-101;
XX Matches 208; Conservative 36; Mismatches 86; Indels 0; Gaps 0;
XX
XX QY 1 MOPLVSLICAYNAEKYFAQSLAAVVGQWNRNLDLIVDGSNDGFTALNRHFOEDGRT 60
XX DB 1 LQPLVSLICAYNAEKYFAQSLAAVVGQWNRNLDLIVDGSNDGFTALNAKDFOKRSRI 60
XX
XX QY RIISNPNLGFIALSLNIGDELAKSGGEXIARTDADIASPGMIKIVGEMEKDRSIIA 120
XX DB 61 KIILAQONSLIPSLNIGDELAKSGGEXIARTDADIASPGMIKIVGEMEKDRSIIA 120
XX
XX QY 121 MGAMLEVLSEENKSVLAIAIRNGATYDKPTRHEDIYAVFPFGNPJHNTMTIMRSYIDG 180
XX DB 121 MGAMLEVLSEENKSVLAIAIRNGATYDKPTRHEDIYAVFPFGNPJHNTMTIMRSYIDG 180
XX
XX QY 181 GLFEPDPAIHAEDYKFMPEYAGKGLRLAYVPEALVYKRFHDDOTSSKYNLQORPTAWKIKE 240
XX DB 181 GLRYNTERDMAEDYQFYWDVSKLGRALAYPEALVYKRLAHNOVSSKSTIQHETIAQGIQK 240
XX
XX QY 241 EIRAGYKKAAGIAGVADCLNYGLKSTAYALYERKALSGDIGCLRFLYFLSLEKYSYL 300
XX DB 241 TARNDFLQSGWGFTRPDSLELYQIKAVAYELLEKHLPEEDFEARARFLYCGFRTDTPPA 300
XX
XX QY 301 TDLIDFLTRVYMRKRLFAAPYRKILKKMLR 330
XX DB 301 GAWLIDFAADGKMRLFTMROYFGILHRLIK 330
XX
XX RESULT 6
XX ID AAM89332 standard; peptide; 50 AA.
XX AC AAM89332;
XX DT 26-FEB-1999 (first entry)
XX
```

```
DE Neisseria gonorrhoeae IgtD C-terminal peptide.
XX
XX KM Neisseria meningitidis; IgtC; IgtB; beta-1,4-galactosyltransferase;
XX KM glycosyltransferase; proteolytic enzyme.
XX
XX OS Neisseria gonorrhoeae.
XX
XX PN WO9854331-A2.
XX PD 03-DEC-1998.
XX
XX PF 26-MAY-1998; 98WO-IB00975.
XX PR 27-MAY-1997; 97US-0047751.
XX
XX PA (CANA ) NAT RES COUNCIL CANADA.
XX PI Wakarchuk WM, Young NM;
XX DR WPI; 1999-035177/03.
XX
XX PT Expressing high levels of glycosyltransferases - comprises use of
XX PT either host cells deficient in proteolytic enzymes or modified
XX PT glycosyltransferase genes deleted in a proteolytic recognition site
XX PS Example 1; Fig 8; 61pp; English.
XX
XX CC A method has been developed of expressing a glycosyltransferase in a
XX CC host cell. The method comprises introducing into the host cell a nucleic
XX CC acid encoding the glycosyltransferase and incubating the host cell under
XX CC conditions appropriate for expression of the glycosyltransferase, where
XX CC the host cell substantially lacks a protease that cleaves polypeptides
XX CC between two consecutive positively charged amino acid residues. The
XX CC glycosyltransferase can be used in in vitro production of
XX CC oligosaccharide structures which are potential therapeutic agents for
XX CC use in the manipulation of cell-cell recognition events, particularly
XX CC adhesion of bacteria and viruses to mammalian cells and leukocyte-
XX CC endothelial cell interaction through selectins in inflammation. The
XX CC method provides more readily recoverable active glycosyltransferases
XX CC than prior art methods involving mammalian glycosyltransferases. The
XX CC present sequence represents a C-terminal peptide from Neisseria
XX CC gonorrhoeae IgtD from the present invention.
XX
XX SQ Sequence 50 AA:
XX
XX Query Match 14.7%; Score 260; DB 20; Length 50;
XX Best Local Similarity 100.0%; Pred. No. 6.2e-20;
XX Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 285 RLFLYEFSLSEKYSLTDLDFLTRVYMRKRLFAAPYRKILKKMLRPWKY 334
XX DB 1 RLFLYEFSLSEKYSLTDLDFLTRVYMRKRLFAAPYRKILKKMLRPWKY 50
XX
XX RESULT 7
XX ID AAB96313 standard; Protein; 298 AA.
XX AC AAB96313;
XX DT 29-OCT-2001 (first entry)
XX
XX DE Putative glycosyltransferase, involved in cell wall biogenesis #1.
XX
XX KM Hyperthermophilic archaeon; hyperthermophilic protein.
XX
XX OS Pyrococcus abyssi.
XX
XX PN FR2792651-A1.
XX DT 27-OCT-2000.
XX PD 21-APR-1999; 99FR-0005034.
XX PF
```

```

XX 21-APR-1999; 99FR-0005034.
PR (CNRS ) CNRS CENT NAT RECH SCI.
PA (IFRE ) IREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX Querellou J, Weissbach J, Saurin W, Hellig R;
XX WPI: 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry -
XX
XX Claim 7; Pages 981-982; 1657pp: French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
XX a hyperthermophilic archaeon, which is isolated from deep-sea
XX hydrothermal vents. The present sequence is one such P. abyssi protein.
XX The proteins of the present invention have various potential industrial
XX uses, since the proteins are stable at very high temperatures, some up to
XX 110 degrees centigrade.
XX Note: This patent is in the same patent family as WO200065062, which
XX contains additional sequences as shown in AAB99132-AB99143,
XX AAH75903-AAH75920 and AAG66436.
XX
XX Sequence 298 AA:
SQ
Query Match 13.8%; Score 243; DB 22; Length 298;
Best Local Similarity 26.3%; Pred. No. 6.1e-17;
Matches 83; Conservative 48; Mismatches 128; Indels 56; Gaps 10;
QY 2 OPLVSVLCAVNAEYFQASIAVVGQTRNDLIIYDGSIDGSPATARHFOEDGRIR 61
DB 3 RIVSVIITPTNRANLRRALASVLSQKFKDFELIVDASTDNPVEVEST--EDGRIR 60
QY 62 IISNPNLGFIALNIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEME--KDRSI 118
DB 61 YIRLKNNSGCFIARNIGT---KKAKGRFIALDDDEWLPRLFEVQVAKFENLGEFGEV 116
QY 119 IAMGAMLEVLSERNKSVLAIAIRNGAIWDK--PTRHEDIIVAFPPGPNHNMTIMRRS 176
DB 117 YVGFGEY-----VSODGRILGKRLPKHKGDIYSHLLKENFISPTLLIRRE 162
QY 177 VIDGIRPDPATIHADYKFWYEAGLGLRAYPPLVYRRHOOTS---SKYNQQR 233
DB 163 CFFKAGLFPRLSSODWMLRIAYYFEDVDETIAYYHGKQISFMKKYIPGRER 222
QY 234 TAWKIKEIRAGYWRKAAGIAGVADCLNYGL-----KSTAYALYEKALSGODIGCLR 285
DB 223 LIRK-----HDIWKPKI-LSIHLSQMGILLLSNNTOKGLKYLTYSTAIAP----- 269
QY 286 LFLYEFLEKYSYL 300
DB 270 -----LNLNENYMI 277

```

```

RESULT 8
AAB47426
ID AAB47426 standard; Protein; 332 AA.
XX
XX AAB47426;
AC
XX
XX 17-OCT-2001 (first entry)
XX
XX Epsn.
XX
XX EPS plasmid; Lactococcus lactis subspecies cremoris Ropy352;
XX exopolysaccharide; EPS352; milk; thickener; glycosyltransferase;
XX biosynthesis; fruit juice.
XX
XX Lactococcus lactis.
OS

```

```

XX WO200157234-A2.
XX
XX 09-AUG-2001.
XX
XX 02-FEB-2001; 2001WO-US03404.
XX
XX 02-FEB-2000; 2000US-0179888.
XX 16-OCT-2000; 2000US-0241098.
XX
XX (USDA ) US DEPT OF AGRICULTURE.
XX
XX Trempey JE, Knoshaug EP, Sandine WE, Ahlgren JA, Dierksen KP;
XX WPI: 2001-488889/53.
XX N-PSDB; AAH43198.
XX
XX New bacterium useful in pharmaceutical formulations, food products and
XX beauty cosmetics, comprises characteristics of Lactococcus lactis
XX cremoris Ropy 352 -
XX
XX Claim 28; Page 67-68; 73pp: English.
XX
XX The sequences given in AAB47426-27 are encoded by a fragment of the EPS
XX plasmid derived from L. lactis subspecies cremoris Ropy352. The EPS
XX plasmid is about 32 kb in size and encodes at least 13 active genes.
XX The enzymes encoded by these genes allow the bacteria to produce an
XX exopolysaccharide, designated EPS352. When EPS352 is expressed in or
XX added to milk, it imparts highly desirable sensory characteristics to
XX the milk, including making the milk very thick, with a very smooth
XX mouth-feel, and slightly sweet with an obvious chewable-bite. Open
XX reading frames (ORF's) M and N show homology to glycosyltransferase
XX involved in EPS352 biosynthesis. L. lactis cremoris Ropy352 is
XX deposited with the USDA-ARS-NCAUR-NRL as deposit accession number
XX NRL-B-30229. EPS352 is useful for thickening a liquid selected from
XX milk, a milk-based liquid, a whey-based liquid, a soy-based liquid,
XX and a fruit-juice. It is also useful as an additive in pharmaceutical
XX products, beauty care products and coating agents.
XX
XX Sequence 332 AA:
SQ
Query Match 12.7%; Score 223.5; DB 22; Length 332;
Best Local Similarity 24.8%; Pred. No. 8.4e-15;
Matches 61; Conservative 55; Mismatches 81; Indels 49; Gaps 8;
QY 4 LYSVILCAVNAEYFQASIAVVGQTRNDLIIYDGSIDGSPATARHFOEDGRIRII 63
DB 5 LISIIVPVYNSERYLAAHLSHSLNLYQNIIEVILLNDGSTDSOELISSFOKKDKRIKLY 64
QY 64 SNPNLGFIALNIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKRSTIAMGA 123
DB 65 -NTRKNGVSHARNYIDR---ASGSYIMFLPDDPDYDKSYCLEMGLINKRNADVMSN 119
QY 124 W-----LE---VLSEENN-KSVLAIAIRNGAIWDKPTRHEDIIVAFPPG 163
DB 120 YIICKGNITPMVNNDLLECEGLSDKTMRSILSDTGKFGFW----- 163
QY 164 NPIHNTMTIMRSVIDGIRPDPATIHADYKFWYEAGLGLRAYPPLVYRRHOOD 222
DB 164 -----TRIFRRNVIN-NVKFESINYLEDMFNISIVHARIATVNRHVFYLOREDS 216
QY 223 TSSKYN 228
DB 217 ASKFS 222

```

```

RESULT 9
AAG90151
ID AAG90151 standard; Protein; 274 AA.
XX
XX AAG90151;
AC
XX
XX

```

DT 26-SEP-2001 (first entry)
XX C glutamicum protein fragment SEQ ID NO: 3905.
XX
XX Coryneform bacterium: amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis.
XX Corynebacterium glutamicum.
OS
XX
XX EP108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
DR N-PSDB: AAH5370.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX
XX Claim 17: SEQ ID NO: 3905; 246pp + Sequence Listing: English.
XX

CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX

Seq Sequence 274 AA;

Query Match 12.4%; Score 219.5; DB 22; Length 274;
Best Local Similarity 27.1%; Pred. No. 1.7e-14;
Matches 78; Conservative 57; Mismatches 108; Indels 45; Gaps 13;

QY 3 PLYSVLICAVNAEKYFQASLAAYVGGTWRNLDLIYDGGSTGTPA-IARHQEQDGR1 60
DB 6 PLISVVIPTIAYEYCSQSIKVCSEONYENWQIVLV---LDCAPIKDVPQWKEHE-RI 60
QY 61 RIISNRNFGFIASINGIDELAKSGGGEVIARTDADDIASPMIKIYEMKKDSIIA 120
DB 61 KIYEQKIRCTPTSLNNGI---KASDGLIARLSDSDLAAPSRLSQOEFLNHPYITC 116
QY 121 MGAMLEVLSEENKSVLAARNGATW---DKPTRHEDIVAVFPNGPIHNMTIMRRS 176
DB 117 VAKTKHINE-----HGKIFGOSADLPT-SQDIRQLLVKNPIIHSVWRKQ 163
QY 177 VID--GGLRPDPAVYIAEDYKFWYEGAKGLRILAYPDAVKYFPHODQISSKYNLQDRT 234
DB 164 VVEOIGYSLE--MTRSOBYELFLRLSIGALIGYLDSELSYSYRHHGQHSRKRKTSPEKYY 221
QY 235 AMKIKBEIR-AGYWAAGTA-----VGADCLNY-GILKS---TAY 269
DB 222 WILKRMELASFLKRSVVRQIFLNFITWGAOVTRYLGLKRAKGFMTGY 269

RESULT 10
AA54072
ID AAY54072 standard; Protein; 316 AA.

XX AAY54072;
XX
XX 27-MAR-2000 (first entry)
DT
XX

DE Enzyme EPS8 which is involved in exopolysaccharide biosynthesis.

XX
XX Exopolysaccharide; EPS; ESP enzyme; EPS1; EPS2; EPS3; EPS4; EPS5;
KM EPS6; EPS7; EPS8; EPS9; EPS10; Streptococcus thermophilus strain SF139;
KM activated D-galactose pyranose; saccharide; beta-glycosyltransferase;
KM transporter; food; fermented milk product; yoghurt; cheese;
KM flavour stability; organoleptic property.
XX
XX

OS Streptococcus thermophilus.

XX
XX MO9962316-A2.
PN

XX 09-DEC-1999.
PD

XX 22-APR-1999; 99WO-EP02841.
XX

XX 22-APR-1998; 98EP-0201310.
XX

XX 22-APR-1998; 98EP-0201311.
XX

XX 22-APR-1998; 98EP-0201312.
XX

XX (NEST) SOC PROD NESTLE SA.
XX

XX Stinglele F, Germond JE, Lamothe G;
XX

XX WPI: 2000-097267/08.
XX

XX N-PSDB: AA445258, AAY54072, AAY54073, AAY54074.
XX

PT New recombinant enzymes for synthesis of exopolysaccharides,
PT particularly in lactic acid bacteria, for improving properties of
PT fermented milk products -
XX

PS Claim 3; Page 105-106; 162pp; French.

XX
XX AAY54065-74 represent enzymes involved in the biosynthesis of
XX exopolysaccharides (EPS). These enzymes are designated EPS1-EPS10, and
XX are encoded by open reading frames eps1-eps10. The enzymes are isolated
XX from Streptococcus thermophilus strain sf139. The proteins are used
XX in a method for the synthesis of EPS, which includes at least one step
XX of forming a bond (alpha or beta-isomer) between C-1 (carrying the
XX reducing aldehyde function, of an activated D-galactose pyranose), and
XX a phosphate on a lipophilic or proteinaceous primer. Stepwise synthesis
XX of EPS occurs with, in each step, addition of a new sugar unit, through
XX its hemi-acetyl function, to an alcoholic hydroxyl of a second sugar
XX unit, present at the end of a chain of sugar residues bonded to the
XX primer. EPS1 to EPS4 are used to elongate the EPS chain, and to regulate
XX EPS production. EPS5 creates new bonds between saccharides, EPS6 is used
XX in the biosynthesis of EPS, EPS7 and EPS8 are beta-glycosyltransferases,
XX EPS9 is a transporter of repetitive units, and EPS10 catalyses the
XX conversion of a pyranose form of a beta-D-galactose to the furanose
XX form. The EPS enzyme are used to improve properties of foods,
XX particularly fermented milk products such as yoghurt and cheese,
XX e.g. their organoleptic properties and flavour stability.
XX

Seq Sequence 316 AA;

Query Match 12.2%; Score 215; DB 21; Length 316;
Best Local Similarity 23.1%; Pred. No. 6.2e-14;
Matches 81; Conservative 59; Mismatches 105; Indels 106; Gaps 14;

QY 1 MOPLSVLICAVNAEKYFQASLAAYVGGTWRNLDLIYDGGSTGTPA-IARHQEQDGR1 60
DB 1 MNPISIIIVPTVWEYIRTCIESIIAQYIRNIEVLIYDGGSTGSLAVIISDLICSHHNI 60

CC PMCS catalyzes glycosaminoglycan polymerisation to produce
CC chondroitin: a linear polysaccharide which has viscoelastic properties
CC which makes it useful for a number of applications. Chondroitin can be
CC used with hyaluronic acid (HA) to coat medical devices e.g. catheters and
CC sensors to reduce tissue abrasion. In addition, they can be used as
CC bioadhesives for haemostatic sealing and healing of wounds and surgical
CC incisions; and as biomaterials that provide sustained delivery of
CC encapsulated drugs, to wounds, ulcers, injuries or surgical sites.

XX Sequence 965 AA;

Query Match 11.8%; Score 209; DB 21; Length 965;

Best Local Similarity 23.7%; Pred. No. 1.5e-12; Indels 60; Gaps 9;

Matches 75; Conservative 53; Mismatches 129;

QY 3 PLVSVLICAVNAEKYFAQSIAAVVGTWRNLDILIVDGSSTGCTPALAHFQDGRIRI 62
DB 433 PLVSVIIPAINCANYIORCVDSALNQTVDLEVCICNDGSTDWTLEYNKLYGNRPVR 492

QY 63 ISNPRNLGFASINIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKRSIIAMG 122
DB 493 MSKPN--GGIASASNAVSPAK--GYIIGQLSDDYLEPDVALELCKEFLKDKTLAC-- 545

QY 123 AMLEVLSEENKSVLAIAIRNGAIPDKPTRHEDIVAVFPFGNPIHNNMTIMRS--VIDG 180
DB 546 ----VTTNNVNPDSGLIANGYNWPEFSREKLTMTMI-----AHFRMTIIRAMHLTDG 596

QY 181 GLRPDPAYTHAEDYKFWYEGKGLRLAYPEALVKYRFHODOTSSKYNLQORRTAMKIKE 240
DB 597 ---FNEINENAVDYDMFLKLEVEGKFKHLNKICYNRVLHGDNTSIRK----- 639

QY 241 EIRAGYWKAGIAGVADCLNLYGLKSTAYALYERKALSGODIGCLRFLYEFYFLSLE---K 297
DB 640 -----KLGIOKKNHFYVNVQSLNRQGI---NYYNQKFDLDESRR 677

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

PT applications, e.g. for eye or joint applications, for moisturizer or
PT wound dressings -
XX
XX
PS Claim 3; Page 119; 125pp; English.

CC The present invention relates to the coding sequence of the Pasteurella
CC multocida chondroitin synthase. A chondroitin polysaccharide may be used
CC as a hyaluronan polysaccharide substitute in medial or cosmetic
CC applications, for example in eye or joint applications, for moisturiser
CC or wound dressings. The enzyme may be used in covalently coupling
CC specific drugs, proteins or toxins to the structurally modified
CC chondroitin for general or targeted drug delivery or radiological
CC procedures, covalently cross linking the hyaluronic acid itself or to
CC other supports to achieve a gel or other three dimensional biomaterial
CC with stronger physical properties, and covalently linking hyaluronic acid
CC to a surface to create a biocompatible film or monolayer. The present
CC sequence is one version of the protein of the invention.

XX Sequence 965 AA;

Query Match 11.8%; Score 209; DB 23; Length 965;

Best Local Similarity 23.7%; Pred. No. 1.5e-12; Indels 60; Gaps 9;

Matches 75; Conservative 53; Mismatches 129;

QY 3 PLVSVLICAVNAEKYFAQSIAAVVGTWRNLDILIVDGSSTGCTPALAHFQDGRIRI 62
DB 433 PLVSVIIPAINCANYIORCVDSALNQTVDLEVCICNDGSTDWTLEYNKLYGNRPVR 492

QY 63 ISNPRNLGFASINIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKRSIIAMG 122
DB 493 MSKPN--GGIASASNAVSPAK--GYIIGQLSDDYLEPDVALELCKEFLKDKTLAC-- 545

QY 123 AMLEVLSEENKSVLAIAIRNGAIPDKPTRHEDIVAVFPFGNPIHNNMTIMRS--VIDG 180
DB 546 ----VTTNNVNPDSGLIANGYNWPEFSREKLTMTMI-----AHFRMTIIRAMHLTDG 596

QY 181 GLRPDPAYTHAEDYKFWYEGKGLRLAYPEALVKYRFHODOTSSKYNLQORRTAMKIKE 240
DB 597 ---FNEINENAVDYDMFLKLEVEGKFKHLNKICYNRVLHGDNTSIRK----- 639

QY 241 EIRAGYWKAGIAGVADCLNLYGLKSTAYALYERKALSGODIGCLRFLYEFYFLSLE---K 297
DB 640 -----KLGIOKKNHFYVNVQSLNRQGI---NYYNQKFDLDESRR 677

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694

QY 298 YSLTDLDFLTDRVMRK 314
DB 678 YIFNKTAEYQEMDMK 694


```
XX DeAngelis P, Kumari K, Weigel PH;
XX
XX WPI: 1999-337486/28.
DR N-PSDB: AAX58857.
XX
XX Nucleic acid encoding hyaluronate synthase for production of
PT hyaluronic acid with controlled molecular weight and targeting
PT specificity
XX
XX Disclousure: Page 123-125; 125pp; English.
PS
XX
XX This present sequence represents the hyaluronate synthase (pmHAS)
CC of Pasteurella multocida Carter Type A. The pmHAS enzyme has
CC different kinetic optima with respect to pH and metal ion
CC dependence, and different Km values compared with the HAS enzymes
CC of Streptococcus equisimilis (see AAY06206) and Streptococcus
CC pyogenes. Km values are about 2- to 3-fold lower for UDP sugars,
CC and Vmax values are about 2- to 3-fold higher. The invention
CC provides recombinant vectors containing hyaluronate synthase DNA,
CC especially S. equisimilis hyaluronate synthase DNA (see AAX58841),
CC and prokaryotic or eukaryotic host cells which produce the enzyme
CC and its hyaluronate acid product, particularly a product with
CC modified structure or molecular size. The hyaluronate acid produced
CC this way is purer than that produced by conventional methods.
XX
XX Sequence 972 AA:
SQ
Query Match 11.8%; Score 209; DB 20; Length 972;
Best Local Similarity 21.8%; Pred. No. 1.5e-12;
Matches 96; Conservative 60; Mismatches 146; Indels 138; Gaps 14;
QY 3 PLYSVLICAYNAEKYPAQSLAAVVGOTWRNDILIVDSDSTGTPAIARHPQDGRIRI 62
DB 440 PLYSVITIPYANCANYIQRCVDSALNOTVVDLEVCIONDSTDTLEVIKLYGNRPVRH 499
QY 63 ISNPNRLGFIASLNLGDELASGSGGEYIARTDADIASPGWIEKIVGEMKDRSIANG 122
DB 500 MSKPN--GGIASASNAVSEFAK--GYTIGQSDSDYLEPDAVELCLKEFLKDKTLAC-- 552
QY 123 AMLEVLSENNKSVLAALIRNGAIPDKPRHEDIYAVFFPGNPIHNNTIMRRS--VIDG 180
DB 553 ---YTTNRNVNPDGSLIANGNMPEFSREKLTAMI---AHFRMFTIRAMHLTDG 603
QY 181 GLRFPDPAIYHAEDYKFWYAGKIGRLAYYPALVKYRFHODQTSK----- 226
DB 604 ---FNKEIENAVDYDFMLKISEVGKFKHLNKCYNRVLHGDNSTIKKLGIOKKNHVVVN 660
QY 227 -----YNLQO-----RRTAW----- 236
DB 661 QSLNRQGITYYNVDDEFDLDESRRKTYFNKTAEQEIEDILKIKTIONKDAKIAVSIFYP 720
QY 237 -----KIKEIRAGYWK--AAGIAGVADCLNY 261
DB 721 NTLNGLVKKLNNIIEYKNKIEYIVLHVDKNHLTPDIKKEHLEFYHKNQVNIILNNDISYY 780
QY 262 ---GLIKSTAYALYEKALSGODIGCL-----RLFL---YEYLSLEKYSLTDLDF 306
DB 781 TSNRLIKTAHLSNINKLSQNLNCEYIIFDNHDSLFAVNDSTAY---MKKIDVGNMESA 837
QY 307 LTRDVRKRLFAAPQYRKILK 326
DB 838 LTHDWIEKINAHBPFRKLIK 857
RESULT 15
AAY43099 standard; Protein; 972 AA.
XX
XX AAY43099;
AC
XX
XX 01-FEB-2000 (first entry)
XX
```

```
DE P. multocida hyaluronate synthase (PmHAS) amino acid sequence.
XX
XX Hyaluronate synthase; PmHAS; hyaluronan; hyaluronic acid; HA; cosmetic;
KW drug delivery; angiogenesis; wound healing; capsule synthesis;
KW fowl cholera; shipping fever.
XX
XX Pasturella multocida.
OS
XX
XX WO951265-A1.
XX
XX 14-OCT-1999.
PD
XX
XX 01-APR-1999; 99WO-US07289.
PF
XX
XX 02-APR-1998; 98US-0080414.
PR
XX
XX 26-OCT-1998; 98US-0178851.
PR
XX
XX (OKLA ) UNIV OKLAHOMA.
PA
XX
XX DeAngelis P;
PI
XX
XX WPI: 2000-013032/01.
DR
XX
XX N-PSDB: AA35589.
DR
XX
XX New isolated hyaluronate synthase nucleic acids, used for the
PT production of hyaluronate, for developing antibiotics and vaccines
PT and for diagnostic applications
XX
XX Claim 70; Fig 20; 121pp; English.
PS
XX
XX This is the Pasturella multocida hyaluronate synthase (PmHAS) amino acid
CC sequence. Hyaluronate acid (HA) or hyaluronan, is a polysaccharide that
CC serves both structural and recognition roles in higher animals. Bacteria
CC produce extracellular capsules of HA which mimic their host HA and aid
CC escape from a host immune response. The invention includes a vector
CC containing the PmHAS nucleotide sequence which can be used for the
CC production of HA. Also, specific changes to the HS coding sequence can
CC result in the production of HA having a modified size distribution or
CC structural configuration and functional properties. The HA products can
CC be used in e.g. drug delivery, angiogenesis and wound healing,
CC stabilisation of recombinant proteins and in cosmetics. The HS nucleic
CC acids can also be used to develop agents to block capsule synthesis by
CC pathogens and act as antibiotics. The avirulent P. multocida strains can
CC be used as vaccines for fowl cholera or shipping fever.
XX
XX Sequence 972 AA:
SQ
Query Match 11.8%; Score 209; DB 21; Length 972;
Best Local Similarity 21.8%; Pred. No. 1.5e-12;
Matches 96; Conservative 60; Mismatches 146; Indels 138; Gaps 14;
QY 3 PLYSVLICAYNAEKYPAQSLAAVVGOTWRNDILIVDSDSTGTPAIARHPQDGRIRI 62
DB 440 PLYSVITIPYANCANYIQRCVDSALNOTVVDLEVCIONDSTDTLEVIKLYGNRPVRH 499
QY 63 ISNPNRLGFIASLNLGDELASGSGGEYIARTDADIASPGWIEKIVGEMKDRSIANG 122
DB 500 MSKPN--GGIASASNAVSEFAK--GYTIGQSDSDYLEPDAVELCLKEFLKDKTLAC-- 552
QY 123 AMLEVLSENNKSVLAALIRNGAIPDKPRHEDIYAVFFPGNPIHNNTIMRRS--VIDG 180
DB 553 ---YTTNRNVNPDGSLIANGNMPEFSREKLTAMI---AHFRMFTIRAMHLTDG 603
QY 181 GLRFPDPAIYHAEDYKFWYAGKIGRLAYYPALVKYRFHODQTSK----- 226
DB 604 ---FNKEIENAVDYDFMLKISEVGKFKHLNKCYNRVLHGDNSTIKKLGIOKKNHVVVN 660
QY 227 -----YNLQO-----RRTAW----- 236
DB 661 QSLNRQGITYYNVDDEFDLDESRRKTYFNKTAEQEIEDILKIKTIONKDAKIAVSIFYP 720
QY 237 -----KIKEIRAGYWK--AAGIAGVADCLNY 261
```

Db	721	NTLNGLYKKLNNIEYKNIFVIVLHVKNHNLPPDIKKEILAFYHKHQNILLNNDISYY	:	780
Qy	262	---GLKSTAYALYEKALSGODIGCL-----RLEL---YEFYLSLEKYSTDLDF		306
Db	781	TSNRLITEAHLNSINKLSQNLNCEYIIFDNHDSLFVKNDSYAY--MKRYDGMNFS		837
Qy	307	LTDVVMRKLFAPQYRKILK		326
Db	838	LTHDWIEKINAHPPFKKLIK		857

Search completed: December 2, 2002, 11:59:02
job time : 35.9785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 12.544 Seconds
(without alignments)
790.458 Million cell updates/sec

Title: US-10-007-267-5

Perfect score: 1764

Sequence: 1 MGPLSVLICAVNAEKYFAQ.....APQYRKILKKMLRPMKYSY 337

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/DackRillest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1764	100.0	337	1	US-08-312-387B-5
2	1764	100.0	337	1	US-08-683-426-5
3	1764	100.0	337	1	US-08-683-458-5
4	1764	100.0	337	2	US-08-878-360-5
5	1764	100.0	337	4	US-09-333-412-5
6	1761	99.8	337	1	US-08-312-387B-12
7	1761	99.8	337	1	US-08-683-426-12
8	1761	99.8	337	1	US-08-683-458-12
9	1761	99.8	337	2	US-08-878-360-12
10	1761	99.8	337	3	US-08-478-140B-5
11	1761	99.8	337	4	US-09-333-412-12
12	1761	99.8	337	4	US-09-338-943-5
13	1049	59.5	348	1	US-08-312-387B-3
14	1049	59.5	348	1	US-08-683-426-3
15	1049	59.5	348	1	US-08-683-458-3
16	1049	59.5	348	2	US-08-878-360-3
17	1049	59.5	348	4	US-08-333-412-3
18	1046	59.3	348	1	US-08-312-387B-11
19	1046	59.3	348	1	US-08-683-426-11
20	1046	59.3	348	1	US-08-683-458-11
21	1046	59.3	348	2	US-08-878-360-11
22	1046	59.3	348	3	US-08-478-140B-3
23	1046	59.3	348	4	US-09-333-412-11
24	1046	59.3	348	4	US-09-338-943-8
25	1046	59.3	348	4	US-09-338-943-8
26	1046	59.3	348	4	US-09-338-943-8
27	209	11.8	965	4	US-09-437-277-3

28	201.5	11.4	702	4	US-09-437-277-1	Sequence 1, Appl1
29	199	11.3	324	1	US-08-597-236-10	Sequence 10, Appl
30	199	11.3	324	1	US-08-746-682A-10	Sequence 10, Appl
31	185.5	10.5	281	4	US-08-961-083-196	Sequence 196, App
32	162	9.2	674	4	US-08-961-083-200	Sequence 200, App
33	152.5	8.6	93	4	US-08-858-207A-521	Sequence 521, App
34	152	8.6	303	4	US-08-961-083-202	Sequence 202, App
35	150	8.5	270	4	US-08-961-083-198	Sequence 198, App
36	138	7.8	358	4	US-09-134-001C-5633	Sequence 5633, Ap
37	131.5	7.5	727	4	US-09-134-001C-4067	Sequence 4067, Ap
38	116.5	6.6	79	4	US-08-961-083-168	Sequence 168, App
39	116	6.6	418	4	US-09-134-001C-4051	Sequence 4051, Ap
40	99.5	5.6	1410	3	US-09-333-409-3	Sequence 3, Appl1
41	99.5	5.6	1410	4	US-09-568-102-3	Sequence 3, Appl1
42	99.5	5.6	1410	4	US-09-567-969-3	Sequence 3, Appl1
43	99.5	5.6	1410	4	US-09-568-480-3	Sequence 3, Appl1
44	99.5	5.6	1410	4	US-09-568-486-3	Sequence 3, Appl1
45	99.5	5.6	1410	4	US-09-568-472-3	Sequence 3, Appl1

ALIGNMENTS

```
RESULT 1
US-08-312-387B-5
; Sequence 5, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-312-387B-5

Query Match      100.0%  Score 1764;  DB 1;  Length 337;
Best Local Similarity 100.0%;  Pred. No. 1.2e-178;
Matches 337;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

OY 1 MGPLSVLICAVNAEKYFAQSLAAVVGOTWRNLDILYVDGSDGTPATARRHQBDDGT 60
|||||
DB 1 MGPLSVLICAVNAEKYFAQSLAAVVGOTWRNLDILYVDGSDGTPATARRHQBDDGT 60
|||||
OY 61 RIISNPNLIGFTASLNLIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 120

```
Db 61 RIISNPNLGFIA SLNIGDELAKSGGGEYIARTDADDIASPGMIKXIYGE MEKDSIIA 120
QY 121 MGAMLEVLSEENKSVLAALARNGAIMDKPTRHEDIIVAVFPFGNP1HNNTMIRRSVIDG 180
Db 121 MGAMLEVLSEENKSVLAALARNGAIMDKPTRHEDIIVAVFPFGNP1HNNTMIRRSVIDG 180
QY 181 GLRFDPAVTHAEDYKFWYEGKIGRLAYYPEALVKYRFHODOTSSKYNLQOORTAKIKE 240
Db 181 GLRFDPAVTHAEDYKFWYEGKIGRLAYYPEALVKYRFHODOTSSKYNLQOORTAKIKE 240
QY 241 EIRAGYKKAAGIAGVADCLNLYGLKSTAVAYLKALSGODIGCLRFLYEFYFLSEKYSL 300
Db 241 EIRAGYKKAAGIAGVADCLNLYGLKSTAVAYLKALSGODIGCLRFLYEFYFLSEKYSL 300
QY 301 TDLLDFLTDVRMKLFAAPQYRKILKMLRPMKYRSY 337
Db 301 TDLLDFLTDVRMKLFAAPQYRKILKMLRPMKYRSY 337
```

RESULT 2

```
US-08-683-426-5
; Sequence 5, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,426
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-0958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ. ID NO.: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-426-5
```

```
Query Match 100.0%; Score 1764; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-178;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MGPLSVSLICAYNAEKYFAQSILAAVVGQGTWRNDILIVDDGSTDGTPTAIRHFOEDGRI 60
Db 1 MGPLSVSLICAYNAEKYFAQSILAAVVGQGTWRNDILIVDDGSTDGTPTAIRHFOEDGRI 60
```

```
QY 61 RIISNPNLGFIA SLNIGDELAKSGGGEYIARTDADDIASPGMIKXIYGE MEKDSIIA 120
Db 61 RIISNPNLGFIA SLNIGDELAKSGGGEYIARTDADDIASPGMIKXIYGE MEKDSIIA 120
QY 121 MGAMLEVLSEENKSVLAALARNGAIMDKPTRHEDIIVAVFPFGNP1HNNTMIRRSVIDG 180
Db 121 MGAMLEVLSEENKSVLAALARNGAIMDKPTRHEDIIVAVFPFGNP1HNNTMIRRSVIDG 180
QY 181 GLRFDPAVTHAEDYKFWYEGKIGRLAYYPEALVKYRFHODOTSSKYNLQOORTAKIKE 240
Db 181 GLRFDPAVTHAEDYKFWYEGKIGRLAYYPEALVKYRFHODOTSSKYNLQOORTAKIKE 240
QY 241 EIRAGYKKAAGIAGVADCLNLYGLKSTAVAYLKALSGODIGCLRFLYEFYFLSEKYSL 300
Db 241 EIRAGYKKAAGIAGVADCLNLYGLKSTAVAYLKALSGODIGCLRFLYEFYFLSEKYSL 300
QY 301 TDLLDFLTDVRMKLFAAPQYRKILKMLRPMKYRSY 337
Db 301 TDLLDFLTDVRMKLFAAPQYRKILKMLRPMKYRSY 337
```

RESULT 3

```
US-08-683-458-5
; Sequence 5, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ. ID NO.: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-458-5
```

```
Query Match 100.0%; Score 1764; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-178;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MGPLSVSLICAYNAEKYFAQSILAAVVGQGTWRNDILIVDDGSTDGTPTAIRHFOEDGRI 60
Db 1 MGPLSVSLICAYNAEKYFAQSILAAVVGQGTWRNDILIVDDGSTDGTPTAIRHFOEDGRI 60
```

OY 61 RIISPRNLGFIASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 120
DB 61 RIISPRNLGFIASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 120
OY 121 MGAMLEVLSEENNSVLAIAIRNGAIMDKPTRHEDIYAVFPFGNPIHNTMIMRSYIDG 180
DB 121 MGAMLEVLSEENNSVLAIAIRNGAIMDKPTRHEDIYAVFPFGNPIHNTMIMRSYIDG 180
OY 181 GLRPDPATIAHEDYKFWYEAGKLGRLAYPPALVYKRRHODOTSSKYMLQORRTAMKIKE 240
DB 181 GLRPDPATIAHEDYKFWYEAGKLGRLAYPPALVYKRRHODOTSSKYMLQORRTAMKIKE 240
OY 241 EIRAGYKAAAGIAGADCLNGLKSTAYALYERKALSGODIGCLRFLFYEFLSLEKYSL 300
DB 241 EIRAGYKAAAGIAGADCLNGLKSTAYALYERKALSGODIGCLRFLFYEFLSLEKYSL 300
OY 301 TDLDLFTDRVMKRLFAAPQYRKILKMLRPMKYRSY 337
DB 301 TDLDLFTDRVMKRLFAAPQYRKILKMLRPMKYRSY 337

RESULT 4
US-08-878-360-5

; Sequence 5, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,360
; FILING DATE: 18-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,426
; FILING DATE:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEFAX: 133521
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-878-360-5

Query Match 100.0%; Score 1764; DB 2; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-178;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MQLPVSLICAYNAEKYFAQSLAAVVGQWTRNLDILYVDSGTGTPAIARHFOEDGRI 60
DB 1 MQLPVSLICAYNAEKYFAQSLAAVVGQWTRNLDILYVDSGTGTPAIARHFOEDGRI 60
OY 61 RIISPRNLGFIASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 120
DB 61 RIISPRNLGFIASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 120
OY 121 MGAMLEVLSEENNSVLAIAIRNGAIMDKPTRHEDIYAVFPFGNPIHNTMIMRSYIDG 180
DB 121 MGAMLEVLSEENNSVLAIAIRNGAIMDKPTRHEDIYAVFPFGNPIHNTMIMRSYIDG 180
OY 181 GLRPDPATIAHEDYKFWYEAGKLGRLAYPPALVYKRRHODOTSSKYMLQORRTAMKIKE 240
DB 181 GLRPDPATIAHEDYKFWYEAGKLGRLAYPPALVYKRRHODOTSSKYMLQORRTAMKIKE 240
OY 241 EIRAGYKAAAGIAGADCLNGLKSTAYALYERKALSGODIGCLRFLFYEFLSLEKYSL 300
DB 241 EIRAGYKAAAGIAGADCLNGLKSTAYALYERKALSGODIGCLRFLFYEFLSLEKYSL 300
OY 301 TDLDLFTDRVMKRLFAAPQYRKILKMLRPMKYRSY 337
DB 301 TDLDLFTDRVMKRLFAAPQYRKILKMLRPMKYRSY 337

RESULT 5
US-09-333-412-5

; Sequence 5, Application US/09333412
; Patent No. 6342382
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/333,412
; FILING DATE: 15-Jun-1999
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: July 7, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEFAX: 133521
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-333-412-5

Query Match 100.0%; Score 1764; DB 4; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-178;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 M0PLVSLICAYNAEKYFAQSILAAVVGOTWRNLDILIVDGSSTDGTPAIAHFQEDGRI 60
|
DB 1 M0PLVSLICAYNAEKYFAQSILAAVVGOTWRNLDILIVDGSSTDGTPAIAHFQEDGRI 60
|
QY 61 RIISNPNRLGFIASLNIIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKRSIIA 120
|
DB 61 RIISNPNRLGFIASLNIIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKRSIIA 120
|
QY 121 MGAMLEVLSEENKNSVLAIAIRNGAIWDPKPTREDIVAVFPFGNPIHNNTMTIMRSVIDG 180
|
DB 121 MGAMLEVLSEENKNSVLAIAIRNGAIWDPKPTREDIVAVFPFGNPIHNNTMTIMRSVIDG 180
|
QY 181 GRFDPAYIHADYKFWYEWAGKLGRLAYYPEALVYRPHODOTSSKYNLQORRTAWKIKE 240
|
DB 181 GRFDPAYIHADYKFWYEWAGKLGRLAYYPEALVYRPHODOTSSKYNLQORRTAWKIKE 240
|
QY 241 EIRAGYWKAGIAGVADCLNGLKSTAVAYALKESGODICLRLFYEFLSLEKYSL 300
|
DB 241 EIRAGYWKAGIAGVADCLNGLKSTAVAYALKESGODICLRLFYEFLSLEKYSL 300
|
QY 301 TDLDLFLDRVVRKLFAPQYRKILKMLRPWKYRSY 337
|
DB 301 TDLDLFLDRVVRKLFAPQYRKILKMLRPWKYRSY 337

RESULT 6

US-08-312-387B-12
; Sequence 12, Application US/08312387B
; Patent No. 554353
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312.387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-312-387B-12

Query Match 99.8%; Score 1761; DB 1; Length 337;
Best Local Similarity 99.7%; Pred. No. 2.4e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 M0PLVSLICAYNAEKYFAQSILAAVVGOTWRNLDILIVDGSSTDGTPAIAHFQEDGRI 60
:|
:

DB 1 I0PLVSLICAYNAEKYFAQSILAAVVGOTWRNLDILIVDGSSTDGTPAIAHFQEDGRI 60
QY 61 RIISNPNRLGFIASLNIIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKRSIIA 120
DB 61 RIISNPNRLGFIASLNIIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKRSIIA 120
QY 121 MGAMLEVLSEENKNSVLAIAIRNGAIWDPKPTREDIVAVFPFGNPIHNNTMTIMRSVIDG 180
DB 121 MGAMLEVLSEENKNSVLAIAIRNGAIWDPKPTREDIVAVFPFGNPIHNNTMTIMRSVIDG 180
QY 181 GRFDPAYIHADYKFWYEWAGKLGRLAYYPEALVYRPHODOTSSKYNLQORRTAWKIKE 240
DB 181 GRFDPAYIHADYKFWYEWAGKLGRLAYYPEALVYRPHODOTSSKYNLQORRTAWKIKE 240
QY 241 EIRAGYWKAGIAGVADCLNGLKSTAVAYALKESGODICLRLFYEFLSLEKYSL 300
DB 241 EIRAGYWKAGIAGVADCLNGLKSTAVAYALKESGODICLRLFYEFLSLEKYSL 300
QY 301 TDLDLFLDRVVRKLFAPQYRKILKMLRPWKYRSY 337
DB 301 TDLDLFLDRVVRKLFAPQYRKILKMLRPWKYRSY 337

RESULT 7

US-08-683-426-12
; Sequence 12, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683.426
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312.387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-683-426-12

Query Match 99.8%; Score 1761; DB 1; Length 337;
Best Local Similarity 99.7%; Pred. No. 2.4e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 M0PLVSLICAYNAEKYFAQSILAAVVGOTWRNLDILIVDGSSTDGTPAIAHFQEDGRI 60
:|
:

```
Db      1 LQPLVSLICAVNAEKYFAQSLAAVVGQTMNDILIVDGGSTDTGTPAIAHHFOEQDGR1 60
QY      61 RIISNPNRGFIASLNLIGIDELAKSGGGEYIARTDADDIASPGMIETKITYGEMEKDQSTIA 120
Db      61 RIISNPNRGFIASLNLIGIDELAKSGGGEYIARTDADDIASPGMIETKITYGEMEKDQSTIA 120
QY      121 MGAWEVLSEENKSYLAIAIRNGAIWMDKPTRHEDIIVAVPEPGNP1HNNTMTIMRSVIDG 180
Db      121 MGAWEVLSEENKSYLAIAIRNGAIWMDKPTRHEDIIVAVPEPGNP1HNNTMTIMRSVIDG 180
QY      181 GLRFPAYIAHEDYKFWYAGKLGRLAYYPEALVYKRFHQDOTSSKYNLQORRTAWKIKE 240
Db      181 GLRFPAYIAHEDYKFWYAGKLGRLAYYPEALVYKRFHQDOTSSKYNLQORRTAWKIKE 240
QY      241 EIRAGYWKAAAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRFLLEYFLSLEKYSL 300
Db      241 EIRAGYWKAAAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRFLLEYFLSLEKYSL 300
QY      301 TDLLDFLTDRVMKRLFAAPQYRKILKMLRPMKYRSY 337
Db      301 TDLLDFLTDRVMKRLFAAPQYRKILKMLRPMKYRSY 337
```

RESULT 8

```
US-08-683-458-12
; Sequence 12, Application US/08683458
; Patent No. 5798233
```

GENERAL INFORMATION:

```
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
```

```
CITY: Hackensack
STATE: New Jersey
```

```
COUNTRY: USA
ZIP: 07601
```

COMPUTER READABLE FORM:

```
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,458
```

FILING DATE:

```
CLASSIFICATION: 435
```

```
PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: 08/312,387
```

```
FILING DATE: September 26, 1994
```

CLASSIFICATION: 435

```
ATTORNEY/AGENT INFORMATION:
```

```
NAME: Jackson Esq., David A.
```

```
REGISTRATION NUMBER: 26,742
```

```
REFERENCE/DOCKET NUMBER: 600-1-095A
```

```
TELECOMMUNICATION INFORMATION:
```

```
TELEPHONE: 201 487-5800
```

```
TELEFAX: 201 343-1684
```

```
INFORMATION FOR SEQ ID NO: 12:
```

```
SEQUENCE CHARACTERISTICS:
```

```
LENGTH: 337 amino acids
```

```
TYPE: amino acid
```

```
TOPOLOGY: linear
```

```
MOLECULE TYPE: protein
```

```
US-08-683-458-12
```

```
Query Match          99.8%; Score 1761; DB 1; Length 337;
Best Local Similarity 99.7%; Pred. No. 2,4e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 MOPVSLICAVNAEKYFAQSLAAVVGQTMNDILIVDGGSTDTGTPAIAHHFOEQDGR1 60
Db      1 LQPLVSLICAVNAEKYFAQSLAAVVGQTMNDILIVDGGSTDTGTPAIAHHFOEQDGR1 60
QY      61 RIISNPNRGFIASLNLIGIDELAKSGGGEYIARTDADDIASPGMIETKITYGEMEKDQSTIA 120
Db      61 RIISNPNRGFIASLNLIGIDELAKSGGGEYIARTDADDIASPGMIETKITYGEMEKDQSTIA 120
QY      121 MGAWEVLSEENKSYLAIAIRNGAIWMDKPTRHEDIIVAVPEPGNP1HNNTMTIMRSVIDG 180
Db      121 MGAWEVLSEENKSYLAIAIRNGAIWMDKPTRHEDIIVAVPEPGNP1HNNTMTIMRSVIDG 180
QY      181 GLRFPAYIAHEDYKFWYAGKLGRLAYYPEALVYKRFHQDOTSSKYNLQORRTAWKIKE 240
Db      181 GLRFPAYIAHEDYKFWYAGKLGRLAYYPEALVYKRFHQDOTSSKYNLQORRTAWKIKE 240
QY      241 EIRAGYWKAAAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRFLLEYFLSLEKYSL 300
Db      241 EIRAGYWKAAAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRFLLEYFLSLEKYSL 300
QY      301 TDLLDFLTDRVMKRLFAAPQYRKILKMLRPMKYRSY 337
Db      301 TDLLDFLTDRVMKRLFAAPQYRKILKMLRPMKYRSY 337
```

RESULT 9

```
US-08-878-360-12
; Sequence 12, Application US/08878360
; Patent No. 5945322
```

GENERAL INFORMATION:

```
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
```

```
CITY: Hackensack
STATE: New Jersey
```

```
COUNTRY: USA
ZIP: 07601
```

COMPUTER READABLE FORM:

```
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
```

FILING DATE:

```
CLASSIFICATION: 435
```

```
PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: 08/683,426
```

```
FILING DATE:
```

```
APPLICATION NUMBER: 08/312,387
```

```
FILING DATE: September 26, 1994
```

CLASSIFICATION: 435

```
ATTORNEY/AGENT INFORMATION:
```

```
NAME: Jackson Esq., David A.
```

```
REGISTRATION NUMBER: 26,742
```

```
REFERENCE/DOCKET NUMBER: 600-1-095B
```

```
TELECOMMUNICATION INFORMATION:
```

```
TELEPHONE: 201 487-5800
```

```
TELEFAX: 201 343-1684
```

```
INFORMATION FOR SEQ ID NO: 12:
```

```
SEQUENCE CHARACTERISTICS:
```

```
LENGTH: 337 amino acids
```

```
TYPE: amino acid
```

```
TOPOLOGY: linear
```

```
MOLECULE TYPE: protein
```

```
US-08-878-360-12
```

```
Query Match          99.8%; Score 1761; DB 2; Length 337;
```

Best Local Similarity 99.7%; Pred. No. 2.4e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLICAVAAEKYFQSLAAVVGQTRNDILIVDGSIDGTPALARRHQEDGR1 60
DB 1 LOPVSVLICAVAAEKYFQSLAAVVGQTRNDILIVDGSIDGTPALARRHQEDGR1 60
QY 61 RITSNPNLGFISLNLIGDELAKSGGGEYIARTDADDIASPGMIKIKYEMKEDSIIA 120
DB 61 RITSNPNLGFISLNLIGDELAKSGGGEYIARTDADDIASPGMIKIKYEMKEDSIIA 120
QY 121 MGAMLEVLSEENKSVLAAILARNGAIWDKPTRHEDIIVAVFPFGNP1HNNTMIRRSVIDG 180
DB 121 MGAMLEVLSEENKSVLAAILARNGAIWDKPTRHEDIIVAVFPFGNP1HNNTMIRRSVIDG 180
QY 181 GLRFDPAYTHAEDYKFWYEGKLGRLAYYPEALVKYRFHODOTSSKYN1QOORTAKIKE 240
DB 181 GLRFDPAYTHAEDYKFWYEGKLGRLAYYPEALVKYRFHODOTSSKYN1QOORTAKIKE 240
QY 241 EIRAGYWKAGIAGVADCLNYGLKSTAYALYKALSGODIGCLRFLYEF1LSLEKYS1 300
DB 241 EIRAGYWKAGIAGVADCLNYGLKSTAYALYKALSGODIGCLRFLYEF1LSLEKYS1 300
QY 301 TDLLDFLTRVMKRLFAAPQYRKILKMLRPWKYRSY 337
DB 301 TDLLDFLTRVMKRLFAAPQYRKILKMLRPWKYRSY 337

RESULT 10

US-08-478-140B-5
; Sequence 5, Application US/08478140B
; Patent No. 6127153

GENERAL INFORMATION:

APPLICANT: JOHNSON, KARL F.
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,140B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi

REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNTE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-478-140B-5

Query Match 99.8%; Score 1761; DB 3; Length 337;
Best Local Similarity 99.7%; Pred. No. 2.4e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLICAVAAEKYFQSLAAVVGQTRNDILIVDGSIDGTPALARRHQEDGR1 60
DB 1 LOPVSVLICAVAAEKYFQSLAAVVGQTRNDILIVDGSIDGTPALARRHQEDGR1 60
QY 61 RITSNPNLGFISLNLIGDELAKSGGGEYIARTDADDIASPGMIKIKYEMKEDSIIA 120
DB 61 RITSNPNLGFISLNLIGDELAKSGGGEYIARTDADDIASPGMIKIKYEMKEDSIIA 120
QY 121 MGAMLEVLSEENKSVLAAILARNGAIWDKPTRHEDIIVAVFPFGNP1HNNTMIRRSVIDG 180
DB 121 MGAMLEVLSEENKSVLAAILARNGAIWDKPTRHEDIIVAVFPFGNP1HNNTMIRRSVIDG 180
QY 181 GLRFDPAYTHAEDYKFWYEGKLGRLAYYPEALVKYRFHODOTSSKYN1QOORTAKIKE 240
DB 181 GLRFDPAYTHAEDYKFWYEGKLGRLAYYPEALVKYRFHODOTSSKYN1QOORTAKIKE 240
QY 241 EIRAGYWKAGIAGVADCLNYGLKSTAYALYKALSGODIGCLRFLYEF1LSLEKYS1 300
DB 241 EIRAGYWKAGIAGVADCLNYGLKSTAYALYKALSGODIGCLRFLYEF1LSLEKYS1 300
QY 301 TDLLDFLTRVMKRLFAAPQYRKILKMLRPWKYRSY 337
DB 301 TDLLDFLTRVMKRLFAAPQYRKILKMLRPWKYRSY 337

RESULT 11

US-09-333-412-12
; Sequence 12, Application US/09333412
; Patent No. 6342382

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 12:

US-09-333-412-12

US-08-312-387B-3

Query Match 59.5%; Score 1049; DB 1; Length 348;

Best Local Similarity 64.5%; Pred. No. 7.3e-103; Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

```
QY 1 MOPVSVLICAVNAEKYFAQSLAAVVGOTWRNLDLIYDDGSTDGTPAARHFOEDGRI 60
    |||||
DB 1 MOPVSVLICAVNEKYFAQSLAAVVGOTWRNLDLIYDDGSTDGTLATAKDEQKRDSDRI 60
    |||||
QY 61 RIISNPNRGFLASNLIGDELAKS--GGGEYIARPDADDIASPGNTEKIVGEMEDRSI 118
    :|:::|
DB 61 KIQAQNSGLPLSLNIGDELAKSGGGGEYIARDADDIASPGNTEKIVGEMEDRSI 120
    |||||
QY 119 IIMGANLEVLSEENKSVLAATARNGAINDKPTRHEDIYAVFPFGNPJINNTMIMRSYI 178
    |||||
DB 121 IIMGANLEVLSEEDKGNRLARHHKHGKIKKPTRHEDIYAFPPFGNPJINNTMIMRSYI 180
    |||||
QY 179 DGLRDPAYIHAEDYKFWYKGLRGLAYYPEALVKYRFHODOTSSKYNLQORRTAMKI 238
    |||||
DB 181 DGLRDPEDYERMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANQVSSKHSVRQHEIAQGI 240
    |||||
QY 239 KEIRAGYKKAAGIAGADCLNGLKSTAYALYEKALSGODIGCLRPLYEYFELSLEY 298
    :|:::|
DB 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAYELPEKDLPEDEFERARFLYOCFKRTDTP 300
    |||||
QY 299 SLTDLDFLTDRYMRKLFAPQYRKILKKMLR 330
    |||||
DB 301 PSGAMLDFAADGMRRLFTLRQYFGILYRLIK 332
    |||||
```

RESULT 14

US-08-683-426-3

; Sequence 3, Application US/08683426

; Patent No. 5705367

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/683,426

; FILING DATE:

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/312,387

; FILING DATE: September 26, 1994

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-0958

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 348 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-683-426-3

Query Match 59.5%; Score 1049; DB 1; Length 348;

Best Local Similarity 64.5%; Pred. No. 7.3e-103; Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

```
QY 1 MOPVSVLICAVNAEKYFAQSLAAVVGOTWRNLDLIYDDGSTDGTPAARHFOEDGRI 60
    |||||
DB 1 MOPVSVLICAVNEKYFAQSLAAVVGOTWRNLDLIYDDGSTDGTLATAKDEQKRDSDRI 60
    |||||
QY 61 RIISNPNRGFLASNLIGDELAKS--GGGEYIARPDADDIASPGNTEKIVGEMEDRSI 118
    :|:::|
DB 61 KIQAQNSGLPLSLNIGDELAKSGGGGEYIARDADDIASPGNTEKIVGEMEDRSI 120
    |||||
QY 119 IIMGANLEVLSEENKSVLAATARNGAINDKPTRHEDIYAVFPFGNPJINNTMIMRSYI 178
    |||||
DB 121 IIMGANLEVLSEEDKGNRLARHHKHGKIKKPTRHEDIYAFPPFGNPJINNTMIMRSYI 180
    |||||
QY 179 DGLRDPAYIHAEDYKFWYKGLRGLAYYPEALVKYRFHODOTSSKYNLQORRTAMKI 238
    |||||
DB 181 DGLRDPEDYERMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANQVSSKHSVRQHEIAQGI 240
    |||||
QY 239 KEIRAGYKKAAGIAGADCLNGLKSTAYALYEKALSGODIGCLRPLYEYFELSLEY 298
    :|:::|
DB 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAYELPEKDLPEDEFERARFLYOCFKRTDTP 300
    |||||
QY 299 SLTDLDFLTDRYMRKLFAPQYRKILKKMLR 330
    |||||
DB 301 PSGAMLDFAADGMRRLFTLRQYFGILYRLIK 332
    |||||
```

RESULT 15

US-08-683-458-3

; Sequence 3, Application US/08683458

; Patent No. 5798233

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/683,458

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/312,387

; FILING DATE: September 26, 1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-095A

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 348 amino acids

; TYPE: amino acid

;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
us-08-683-458-3

Query Match 59.5%; Score 1049; DB 1; Length 348;

Best Local Similarity 64.5%; Pred. No. 7.3e-103;
Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

```

QY      1 MOPVSVLICAYNAEKYFAOSLAAYVGGOTWRNLDTLIYDDGSTDGTPAIRHFQODGRI 60
Db      1 MOPVSVLICAYNVEKYFAOSLAAYVNOTWRNLDTLIYDDGSTDGTLAIKDFQKRSRI 60
      61 RIISNPNRLGFTASINIGLDELAKS--GGGEYIARTDADDIASPGMIKIVGEMEKDRSI 118
      61 KIIAQONSGILIPSLINIGLDELAKSGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
QY      119 IAMGAMLEVLSEENKSVLAIAIRNGAIDWKPTRHEDIYAVFPFGNPPIHNTMTIMRSYI 178
Db      121 IAMGAMLEVLSEEDGNRLIARHKGKIMKPTRHEDIAAFPPGPNPIHNTMTIMRSYI 180
QY      179 DGGLRFPDPAIYHAEDYKFWYEAGKLGRLAYYPEALVYKRFHODTSSKYNLQORRTAMKI 238
Db      181 DGGLRFDTEROMAEEDYQFWYDYSKLGRLAYYPEALVYKRLHANQVSSKHSVQHETIAQGI 240
QY      239 KEETIRAGYWKAAAGIAYGADCLNYGLKSTAYALYEKALSGODIGCLRFLYEYFLSLEKY 298
Db      241 OKTARNDFLOSMGFETRFDSELEYRQTKAAAYELPEKDLPEEDFERARRLVQCFFKRTDTP 300
QY      299 SITDILDFLTDRVMRKLFAAPOYRKILKKMLR 330
Db      301 PSGAWIDFAADGRMRRLFTLRQYFGILYRLIK 332

```

Search completed: December 2, 2002, 12:03:15
Job time : 13.544 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 12:00:15 ; Search time 6.75447 Seconds
(without alignments)
794.504 Million cell updates/sec

Title: US-10-007-267-5

Perfect score: 1764
Sequence: 1 MQLPLVSLICAYNAEYFAQ.....APQYRKILKMLRPWKYRSY 337

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCr_NEM_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCrUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US09_NEM_PUB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1764	100.0	337	12	US-10-007-267-5
2	1761	99.8	337	12	US-10-007-267-12
3	1046	59.5	348	12	US-10-007-267-3
4	1046	59.3	348	12	US-10-007-267-11
5	218	12.4	150	10	US-09-924-358-29
6	209	11.8	972	9	US-09-879-959-10
7	208	11.8	332	10	US-09-767-041-22
8	207.5	11.8	278	10	US-09-767-041-36
9	194	11.0	297	10	US-09-816-028A-31
10	188	10.7	322	10	US-09-767-041-35
11	187.5	10.6	322	10	US-09-767-041-34
12	187	10.6	301	10	US-09-816-028A-27
13	185.5	10.5	120	10	US-09-767-041-51
14	185.5	10.5	270	10	US-09-816-028A-39
15	185.5	10.5	281	10	US-09-767-272-196
16	185.5	10.5	332	10	US-09-767-041-21
17	176.5	10.0	313	10	US-09-900-038A-1
18	174	9.9	210	10	US-09-767-041-47
19	162	9.2	303	10	US-09-816-028A-29

20	162	9.2	674	10	US-09-765-272-200	Sequence 200, App
21	152	8.6	120	10	US-09-767-041-52	Sequence 52, Appl
22	152	8.6	303	10	US-09-765-272-202	Sequence 202, App
23	152	8.6	706	10	US-09-815-242-4950	Sequence 4950, App
24	152	8.6	715	10	US-09-815-242-10511	Sequence 10511, App
25	150	8.5	270	10	US-09-765-272-198	Sequence 198, App
26	147.5	8.4	187	9	US-09-973-457-4	Sequence 4, Appl1
27	147.5	8.4	187	10	US-09-815-028-7	Sequence 7, Appl1
28	147.5	8.4	187	12	US-10-074-527-4	Sequence 4, Appl1
29	145	8.2	389	10	US-09-816-028A-34	Sequence 34, Appl1
30	127	7.2	256	10	US-09-925-301-883	Sequence 883, App
31	123	7.0	608	10	US-09-924-358-8	Sequence 8, Appl1
32	116.5	6.6	79	10	US-09-765-272-168	Sequence 168, App
33	116.5	6.6	269	10	US-09-767-041-41	Sequence 41, Appl
34	108	6.1	418	10	US-09-816-028A-33	Sequence 33, Appl
35	105	6.0	358	10	US-09-815-242-5714	Sequence 5714, App
36	105	6.0	573	10	US-09-815-242-12474	Sequence 12474, A
37	103	5.8	612	12	US-10-001-851-25	Sequence 25, Appl
38	98	5.6	324	9	US-09-981-876-181	Sequence 181, App
39	95	5.4	306	10	US-09-912-020-275	Sequence 275, App
40	95	5.4	639	10	US-09-789-417-2	Sequence 2, Appl1
41	95	5.4	639	10	US-09-792-451-2	Sequence 2, Appl1
42	95	5.4	639	12	US-10-052-586-364	Sequence 364, App
43	94.5	5.4	626	12	US-10-001-851-27	Sequence 27, Appl
44	93	5.3	844	9	US-10-108-605-185	Sequence 185, App
45	92	5.2	963	10	US-09-801-368-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1

US-10-007-267-5

Sequence 5, Application US/10007267

Patent No. US20020127682A1

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSSEE: Klauber & Jackson

SPREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,267

FILING DATE: 03-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412

FILING DATE: 15-Jun-1999

APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-095

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 337 amino acids

TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-007-267-5

Query Match 100.0%; Score 1764; DB 12; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.7e-166;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLVSLICAVNAEKYFAQSLAAVVGQTMRNLDILIVDGSYDGPALAHROEDGR 60
DB 1 MQLVSLICAVNAEKYFAQSLAAVVGQTMRNLDILIVDGSYDGPALAHROEDGR 60
QY 61 RIISNPNLGFASLNLGDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSIIA 120
DB 61 RIISNPNLGFASLNLGDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSIIA 120
QY 121 MGAWEVLSSENNKSVLAALARNGAIMDKPTRHEDIYAVFPFGNPJHNTMIRRSYIDG 180
DB 121 MGAWEVLSSENNKSVLAALARNGAIMDKPTRHEDIYAVFPFGNPJHNTMIRRSYIDG 180
QY 181 GLRFPAYTHAEDYKFWYEGKLGRLAYYPEALVKYRFHODQTSKYNLQORTANKIKE 240
DB 181 GLRFPAYTHAEDYKFWYEGKLGRLAYYPEALVKYRFHODQTSKYNLQORTANKIKE 240
QY 241 EIRAGYKKAAGIAGVADCLNYGLKSTAYALYKALSGODIGCLRFLYEYFLSEKYSYL 300
DB 241 EIRAGYKKAAGIAGVADCLNYGLKSTAYALYKALSGODIGCLRFLYEYFLSEKYSYL 300
QY 301 TDLIDFLTDVRMKRLFAAPQYRKILKKMLRPWKYRSY 337
DB 301 TDLIDFLTDVRMKRLFAAPQYRKILKKMLRPWKYRSY 337

RESULT 2
US-10-007-267-12

Sequence 12, Application US/10007267
Patent No. US20020127682A1

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,267

FILING DATE: 03-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412

FILING DATE: 15-Jun-1999

APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION/DOCKET NUMBER: 26,742

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-007-267-12

Query Match 99.8%; Score 1761; DB 12; Length 337;
Best Local Similarity 99.7%; Pred. No. 3.4e-166;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLVSLICAVNAEKYFAQSLAAVVGQTMRNLDILIVDGSYDGPALAHROEDGR 60
DB 1 MQLVSLICAVNAEKYFAQSLAAVVGQTMRNLDILIVDGSYDGPALAHROEDGR 60
QY 61 RIISNPNLGFASLNLGDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSIIA 120
DB 61 RIISNPNLGFASLNLGDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSIIA 120
QY 121 MGAWEVLSSENNKSVLAALARNGAIMDKPTRHEDIYAVFPFGNPJHNTMIRRSYIDG 180
DB 121 MGAWEVLSSENNKSVLAALARNGAIMDKPTRHEDIYAVFPFGNPJHNTMIRRSYIDG 180
QY 181 GLRFPAYTHAEDYKFWYEGKLGRLAYYPEALVKYRFHODQTSKYNLQORTANKIKE 240
DB 181 GLRFPAYTHAEDYKFWYEGKLGRLAYYPEALVKYRFHODQTSKYNLQORTANKIKE 240
QY 241 EIRAGYKKAAGIAGVADCLNYGLKSTAYALYKALSGODIGCLRFLYEYFLSEKYSYL 300
DB 241 EIRAGYKKAAGIAGVADCLNYGLKSTAYALYKALSGODIGCLRFLYEYFLSEKYSYL 300
QY 301 TDLIDFLTDVRMKRLFAAPQYRKILKKMLRPWKYRSY 337
DB 301 TDLIDFLTDVRMKRLFAAPQYRKILKKMLRPWKYRSY 337

RESULT 3
US-10-007-267-3

Sequence 3, Application US/10007267
Patent No. US20020127682A1

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,267

FILING DATE: 03-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412

FILING DATE: 15-Jun-1999

APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION/DOCKET NUMBER: 26,742

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 12:

```

:
: TELEFAX: 201 343-1684
:
: TELEPHONE: 201 487-5800
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 348 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-007-267-3

Query Match          59.5%; Score 1049; DB 12; Length 348;
Best Local Similarity 64.5%; Pred. No. 6,8e-96;
Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

QY 1 MOPVSVLICAVNAEKYFAQSLSAAVVGOTWRNDILIVDGSYDGTPTAARHFOEDGR 60
   1 MOPVSVLICAVNVEKYFAQSLSAAVVGOTWRNDILIVDGSYDGTPTAARHFOEDGR 60
DB 1 MOPVSVLICAVNVEKYFAQSLSAAVVGOTWRNDILIVDGSYDGTPTAARHFOEDGR 60
QY 61 RIISNPNLGFIALSLNIGDELAKS--GGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 118
   61 RIISNPNLGFIALSLNIGDELAKS--GGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 118
DB 61 KILAQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
QY 119 IANGAMLEVLSEENKSVLAARNGAINDKPTRHEDIYAVFPFGNPINNTMIMRSYI 178
   119 IANGAMLEVLSEENKSVLAARNGAINDKPTRHEDIYAVFPFGNPINNTMIMRSYI 178
DB 121 IANGAMLEVLSEENKSVLAARNGAINDKPTRHEDIYAVFPFGNPINNTMIMRSYI 180
QY 179 DGLRFPDAIHAEDYKFWYEAQKLGRLAYYPALVYKYPFHODTSKYNLQORRTAWKI 238
   179 DGLRFPDAIHAEDYKFWYEAQKLGRLAYYPALVYKYPFHODTSKYNLQORRTAWKI 238
DB 181 DGLRFPDAIHAEDYKFWYEAQKLGRLAYYPALVYKYPFHODTSKYNLQORRTAWKI 240
QY 239 KESIRAGYKWAAGIAGADCLNGLKSTAYALYKALSGODIGCLRFLYEYFLSLEKY 298
   239 KESIRAGYKWAAGIAGADCLNGLKSTAYALYKALSGODIGCLRFLYEYFLSLEKY 298
DB 241 QKTAARNDFLOSMEKFTREFDLEKRTKAAYELPERKDLPEDEFRARRFLYOCFKRTDTP 300
QY 299 SLTDLDELTDVYRKLFAPQYRKILKKMLR 330
   299 SLTDLDELTDVYRKLFAPQYRKILKKMLR 330
DB 301 PSGAMLDFAADGRMRRLFTLQYFGILYRLIK 332

RESULT 4
US-10-007-267-11
: Sequence 11, Application US/10007267
: Patent No. US20020127682A1
: GENERAL INFORMATION:
:   APPLICANT: Gotschlich, Emil C.
:   TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
:   OLIGOSACCHARIDES, AND GENES ENCODING THEM
:   NUMBER OF SEQUENCES: 12
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Klauber & Jackson
:   STREET: 411 Hackensack Avenue
:   CITY: Hackensack
:   STATE: New Jersey
:   COUNTRY: USA
:   ZIP: 07601
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION NUMBER: US/10/007,267
: FILING DATE: 03-Dec-2001
: CLASSIFICATION: <Unknown>
: PRIORITY INFORMATION:
:   APPLICATION NUMBER: US/09/333,412
:   FILING DATE: 15-Jun-1999
:   APPLICATION NUMBER: 08/312,387
:   FILING DATE: July 7, 1994
: ATTORNEY/AGENT INFORMATION:
:   NAME: Jackson Esq., David A.
:   REGISTRATION NUMBER: 26,742
```

```

:
: REFERENCE/DOCKET NUMBER: 600-1-095
:
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 201 487-5800
:   TELEFAX: 201 343-1684
:   TELEEX: 133521
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 348 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-007-267-11

Query Match          59.3%; Score 1046; DB 12; Length 348;
Best Local Similarity 64.2%; Pred. No. 1.4e-95;
Matches 213; Conservative 34; Mismatches 83; Indels 2; Gaps 1;

QY 1 MOPVSVLICAVNAEKYFAQSLSAAVVGOTWRNDILIVDGSYDGTPTAARHFOEDGR 60
   1 MOPVSVLICAVNVEKYFAQSLSAAVVGOTWRNDILIVDGSYDGTPTAARHFOEDGR 60
DB 1 MOPVSVLICAVNVEKYFAQSLSAAVVGOTWRNDILIVDGSYDGTPTAARHFOEDGR 60
QY 61 RIISNPNLGFIALSLNIGDELAKS--GGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 118
   61 RIISNPNLGFIALSLNIGDELAKS--GGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 118
DB 61 KILAQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
QY 119 IANGAMLEVLSEENKSVLAARNGAINDKPTRHEDIYAVFPFGNPINNTMIMRSYI 178
   119 IANGAMLEVLSEENKSVLAARNGAINDKPTRHEDIYAVFPFGNPINNTMIMRSYI 178
DB 121 IANGAMLEVLSEENKSVLAARNGAINDKPTRHEDIYAVFPFGNPINNTMIMRSYI 180
QY 179 DGLRFPDAIHAEDYKFWYEAQKLGRLAYYPALVYKYPFHODTSKYNLQORRTAWKI 238
   179 DGLRFPDAIHAEDYKFWYEAQKLGRLAYYPALVYKYPFHODTSKYNLQORRTAWKI 238
DB 181 DGLRFPDAIHAEDYKFWYEAQKLGRLAYYPALVYKYPFHODTSKYNLQORRTAWKI 240
QY 239 KESIRAGYKWAAGIAGADCLNGLKSTAYALYKALSGODIGCLRFLYEYFLSLEKY 298
   239 KESIRAGYKWAAGIAGADCLNGLKSTAYALYKALSGODIGCLRFLYEYFLSLEKY 298
DB 241 QKTAARNDFLOSMEKFTREFDLEKRTKAAYELPERKDLPEDEFRARRFLYOCFKRTDTP 300
QY 299 SLTDLDELTDVYRKLFAPQYRKILKKMLR 330
   299 SLTDLDELTDVYRKLFAPQYRKILKKMLR 330
DB 301 PSGAMLDFAADGRMRRLFTLQYFGILYRLIK 332

RESULT 5
US-09-924-358-29
: Sequence 29, Application US/09924358
: Patent No. US20020107376A1
: GENERAL INFORMATION:
:   APPLICANT: Millennium Pharmaceuticals, Inc.
:   TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE FAMILY MEMBERS AND USES THEREFOR
:   TITLE OF INVENTION: 26199, 33530, 33949, 47148, 50226, AND
:   TITLE OF INVENTION: 58764,
:   FILE REFERENCE: 38155-20034.00
:   CURRENT APPLICATION NUMBER: US/09/924,358
:   CURRENT FILING DATE: 2001-08-06
:   PRIOR APPLICATION NUMBER: US 60/229,300
:   PRIOR FILING DATE: 2000-09-01
:   NUMBER OF SEQ ID NOS: 52
:   SOFTWARE: FastSeq for Windows Version 4.0
:   SEQ ID NO 29
:   LENGTH: 150
:   TYPE: PRT
:   ORGANISM: Artificial Sequence
:   FEATURE:
:   OTHER INFORMATION: Consensus amino acid
US-09-924-358-29

Query Match          12.4%; Score 218; DB 10; Length 150;
Best Local Similarity 40.0%; Pred. No. 2.2e-14;
Matches 52; Conservative 26; Mismatches 44; Indels 8; Gaps 5;
```



```

1  PRIOR FILLING DATE: 1999-07-19
2  PRIOR APPLICATION NUMBER: EP98020465.5
3  PRIOR FILLING DATE: 1998-07-22
4  PRIOR APPLICATION NUMBER: EP98020467.1
5  PRIOR FILLING DATE: 1998-07-22
6  NUMBER OF SEQ ID NOS: 53
7  SOFTWARE: PatentIn version 3.0
8  SEQ ID NO 36
9  LENGTH: 278
10 TYPE: PRT
11 ORGANISM: Streptococcus suis
12 FEATURE:
13 NAME/KEY: misc_feature
14 OTHER INFORMATION: CPSiK
15 OS-09-767-041-36

```

Query Match	11.8%	Score 207.5	DB 10	Length 278
Best Local Similarly	25.28%	Pred. No. 5.8e-13		
Matches 62; Conservative	56;	Mismatches 73;	Indels 55;	Gaps 12

[illegible]

```

RESULT 9
US-09-816-028A-31
Sequence 31, Application US/09816028A
Patent No. US20020042369A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarchuk, Warren W.
APPLICANT: National Research Council of Canada
TITLE OF INVENTION: Gangliosides for Biosynthesis of
FILE REFERENCE: 019633-00011105
CURRENT APPLICATION NUMBER: US/09/816,028A
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/455,406
NUMBER OF SEQ ID NOS: 49
SOFTWARE: SeqIdN Ver. 2.1
SEQ ID NO 31
LENGTH: 297
TYPE: PRF
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-09-816-028A-31

```

Query Match	11.0%;	Score 194;	DB 10;	Length 297;
Best Local Similarity	21.8%;	Pred. No. 1.4e-11;		
Matches	81;	Conservative	56;	Mismatches 108; Indels 126; Gaps 14,

QY 5 VSLICAYNAEKYFAQSLAVVGQ^TWRNLDILIVDDGSTDGTPAIARHFQEQDGRIRIS 64

```

Db      4 :|::: ||::: ||::: : |:::|:::|::: |::: : |::: : :|:::|:::|
  ISILPTYNWQIYARALAESCIINTQFKIEIIVVDDGSKSDIYVEYAKKDRKIITH 63

Qy      65 NPNRNFIASINIGLDELAKSGGGEYIARDTDADIASPGWIEKIVGEMERDSIIANGAW 124
  ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
Db      64 NBNENIKLARREYECY-----KVANSPYIMFLDPD-----Y 94

Qy      125 LEVISEENNKSVLAIAIRNANGAIWDKPTREHEDIVAVFPFGPFIHNNTIMRSDYDGLRF 184
  ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
Db      95 LELNACECMKIL-----KNNDIDLLFNAFYLENNKNIIEK-----LNF 134

Qy      185 -DPAYIAHEDYK-----FW-----YEAGKLGK---LAYP- 210
  :|::: :|::: :|::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
Db      135 QEKCYVKRDLFKELLTKNLPMFWMAKVIKKEYLAKAVGLSISENARKINNAEDVILYYPL 194

Qy      211 -----EALVKYRFHODOTSSKYNLOQRTAMKIKEEIRIAGYKKAAGIANGADCL 259
  :|::: :|::: :|::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
Db      195 INISNTIFHLSKNLNYQIINNFSITKTLTLLONIKT--NIOEODVLY- 239

Qy      260 NGGLKASVAYALYERAKLSGQIGCLRFLFYRYEFLSEKYSITDLDPLTDYRWKRLFPAP 319
  ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
Db      240 --LLKKQOY-----NYNENFTLLKILEYFLPIEKYSLSKRNVLCEKI--NIF-- 283

Qy      320 QYRKILKMLR 330

Db      284 -FKKIQFEFYR 293

```

```

RESULT 10
US-09-767-041-35
; Sequence 35, Application US/09767041
; Patent No. US20020055168A1
;
GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4776
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: FCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 35
; LENGTH: 322
; TYPE: PRF
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: CDS1J
US-09-767-041-35

```

Query Match	10.7%;	Score 188;	DB 10;	Length 322;
Best Local Similarity	23.4%;	Pred. No. 6e-11;		
Matches	81;	Conservative	55;	Mismatches 132;
			Indels	78;
			Gaps	15;

```

0Y 5 VLLICAVNAEKYFQSLAAVVGOMWRLDILIYDGSSTQSPAIAPHQEOGRIRI-- 62
   1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1:::
Db 4 ISVIVPVYINVDYTLSSCIESIINOMYKNEILILDDGSVDSDAICKEY-EKDRVATFF 62
   1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1:::
0Y 63 ----ISNERNLGFYASLNGIDELAKSGGGEYIARTDADDIASPGMIKIVGEMEKDRS 117
   1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1:::
Db 63 TMSVSNAARNHGI-----KSTAEYIMFVSDVDVDSRIYERKLYEMIIKRS 110
   1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1:::
0Y 118 IIAAGAMLEVISEENKSVLAIAIRNGAIIMPKPRHEDIYAVPEPGPIHNNIMKRSV 177
   1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1:::
Db 111 DLS-CCLYATFSENIN-----FEVNNP--NIDEAINTV 142
   1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1:::
0Y 178 ID-GGLRFDPAVIAHEDYKFWYEGAKLGRLAYPEALVKYRFHODOTSRS---YNLOQR 233
   1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1::: 1:::

```


Db 4 VSLIPIFNTERYLARECLDSIIISQSYTNLEILLIDGSSSDSSTDLCELYAEDGRKILR 63
QY 63 -----ISNPNLGFASINIGDELAKSGGGEYIARTDADIASPGWIKIYGEKDRS 117
Db 64 LKNGVSNRNGI-----KNSNTANIMFVSDSDIDVGNIVESLYTCLKENDS 111
QY 118 IAMG 122
Db 112 DLSCG 116

RESULT 14

US-09-816-028A-39
; Sequence 39, Application US/09816028A
; Patent No. US20020042369A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 39
; LENGTH: 270
; TYPE: PRP
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: glycosyltransferase from C. jejuni OH4384 (ORF 12a)
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-09-816-028A-39

Query Match 10.5%; Score 185.5; DB 10; Length 270;
Best Local Similarity 21.3%; Pred. No. 8.3e-11;
Matches 65; Conservative 48; Mismatches 79; Indels 113; Gaps 10;

QY 3 PLVSYLICAYNAEKYPAQSLAAVVGQTMRLDILIVDGSSTDPATIAHFOEDGRIRI 62
Db 2 PQSLIILPLFNSCDFISRLQSCINOTLMDIELIIDDKSKNSLWVLEFAKDRIRI 61
QY 63 ISNPNLGFASINIGDELANS GGGEYIARTDADIASPGWIEKIVGEMEKDRSIIAMG 122
Db 62 FQNEMLGTFASNLGV-----LHSSSDFIPLDSDDFLPPDACEIAFKEMKKGFDLLC-- 115
QY 123 AMEVLSENNKSVLAIAIRNGAIWPKTRHEDIVAVFFPGNIHNTIMRSVVDGL 182
Db 116 ----- 115
QY 183 REDPAIHAEDYKFWYEAQKGLRLAYVPEALVYRPHODOTS-----KYNQORRTAMK 237
Db 116 -ED-ATVHVYKTKQF-----YRKKQDEVFQKQKLEFLSKQRHFCMS 155
QY 238 IKEEIRAGIWKAGIAGVADCLWYG-LKSTAYALYKALS-GQDIGCLRLFLYEFLSL 295
Db 156 V-----W-----AKCFKKDIIILKSFEKIKIDERLNGEDV-----LFCYIYFMFC 195
QY 296 EKYSL 300
Db 196 EKIAV 200

RESULT 15
US-09-765-272-196
; Sequence 196, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 196:
SEQUENCE CHARACTERISTICS:
LENGTH: 281 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 196:
US-09-765-272-196

Query Match 10.5%; Score 185.5; DB 10; Length 281;
Best Local Similarity 23.1%; Pred. No. 8.7e-11;
Matches 75; Conservative 53; Mismatches 118; Indels 79; Gaps 13;

QY 25 VVGQTFNRNLDILIVDGSSTDPATIAHFOEDGRIRIISNPNLGFASINIGDELAK 84
Db 5 ILKQYQNIETIIVDGSSTDSNGEIDAFPMQDNRRVRLHOKNGGAQAQKMGIT-SYAK 63
QY 85 SGGGEYIARTDADIASPGWIEKIVGEMEKDRSIIAMGAMLEVLSENN----- 133
Db 64 ---GEYITVSDSDIVKENMIEFLYQVQEKDADVYIGNVYVDESDGNFYFYVGQDFC 120
QY 134 ---KSVLAIAIRNGAIWPKTRHEDIVAVF-----PFGPIHNTIMRSVVDGLRD 185
Db 121 VELLAQEIIMNQAGDW-KFNSSAFILPFLKIKKLEFNEVH-----FSNRRRD 169
QY 186 PAYIHAEDYKFWYEAQKGLRLAYVPEALVYRPHODOTSSTKYNLQORRTAMKIEIRAG 245
Db 170 D---EATMRIFILASK---IYFINDNLVLR-----RRGSIIMRIFEDS 209
QY 246 YKMAAGIAGV---ADCLNVGLKSTAYALYKALS-GQDIGCLRL---FLYEFLSLEK 297
Db 210 -WARDIVEFSKISDCV-----LAGLDVSVLRIRFVNLADVYQOTLEY 252
QY 298 YSLTDLDFLTBRVAKKLEFAPOYR 322
Db 253 HOLITREYKIDICFRLKLEFDAEOR 277

Search completed: December 2, 2002, 12:16:02
Job time : 7.75447 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 14.9563 seconds
(without alignments)
2166.126 Million cell updates/sec

Title: US-10-007-267-5
Perfect score: 1764
Sequence: 1 MQLVSVLICAVNAEKYFAQ.....APQYRKILKMLRPMWKYRSY 337

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR-73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1074.5	60.9	333	1 S70813	glycosyl transferase
2	1069.5	60.6	349	2 D81027	lacto-N-neotetraos
3	1039	58.9	346	2 H81970	lacto-N-neotetraos
4	682.5	38.7	323	1 H64130	glycosyl transferase
5	346	19.6	367	2 G95948	probable glycosylt
6	282.5	16.0	333	2 B97168	glycosyltransferase
7	263.5	14.9	336	2 A97168	glycosyltransferase
8	258.5	14.7	340	2 T44330	glycosyl transferase
9	258	14.6	333	2 H97167	glycosyltransferase
10	249	14.1	271	2 B84114	exopolysaccharide
11	248.5	14.1	290	2 H64431	glycosyl transferase
12	244.5	13.9	343	2 A12091	glucosyltransferase
13	243	13.8	298	2 B75096	glycosyl transferase
14	234	13.3	337	2 AG1920	glycosyl transferase
15	231	13.1	278	2 D70036	hypothetical prote
16	226.5	12.8	301	2 F95205	exopolysaccharide
17	225.5	12.8	318	2 AG2189	glycosyl transferase
18	225	12.8	318	2 AH2189	hypothetical prote
19	225	12.8	334	1 G71153	hypothetical prote
20	224	12.7	732	2 AB4107	hypothetical prote
21	223.5	12.7	311	2 T00087	glycosyltransferase
22	223	12.6	321	2 AG2188	thiamosyltransferase
23	223	12.6	333	2 AH2026	hypothetical prote
24	221	12.5	344	2 G70036	spore coat polysac
25	220	12.5	315	2 T44648	glycosyl transferase
26	219.5	12.4	294	2 E83022	probable glycosyl
27	219.5	12.4	324	2 A69290	probable glycosyl
28	218.5	12.4	322	2 T44647	glycosyl transferase
29	216.5	12.3	323	2 AD2189	hypothetical prote

30	212.5	12.0	344	2 AC0974	probable glycosylt
31	212	12.0	337	2 AD3614	glycosyl transferase
32	211.5	12.0	257	2 EA4107	telchuronic acid b
33	209	11.8	250	2 A64059	glycosyl transferase
34	209	11.8	972	2 T09595	glucuronosyltransf
35	207.5	11.8	281	2 G97777	glycosyl transferase
36	207	11.7	604	2 E97757	hypothetical prote
37	206.5	11.7	318	2 T50039	beta-1,4-galactosy
38	205.5	11.6	318	1 E71630	minor telchuric aci
39	205	11.6	328	2 F98024	raffinose-raffinof
40	204.5	11.6	324	2 AB2190	hypothetical prote
41	203.5	11.5	280	2 E71703	glycosyl transferase
42	203.5	11.5	327	2 AB1211	glycosyltransferase
43	202	11.5	330	2 AH2188	hypothetical prote
44	201	11.4	299	2 B83557	probable glycosyl
45	199	11.3	328	2 F95158	glycosyl transferase

ALIGNMENTS

RESULT 1
S70813
glycosyl transferase A (EC 2.4.-.-) - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S70813
R:Jennings, M.P.; Hood, D.W.; Peak, I.R.A.; Virji, M.; Moxon, E.R.
Mol. Microbiol. 18, 729-740, 1995
A:Title: Molecular analysis of a locus for the biosynthesis and phase-variable expres
A:Reference number: S70812; MUID:96414473; PMID:8817494
A:Accession: S70813
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-333 <JEN>
A:Cross-references: EMBL:U25839; NID:9973183; PIDN:AAC44084.1; PID:9973185
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics:
A:Gene: lgtA
A:Start codon: TTG
C:Superfamily: Neisseria meningitidis glycosyl transferase A
C:Keywords: glycosyltransferase

Query Match	60.9%	Score 1074.5	DB 1	Length 333
Best Local Similarity	65.9%	Pred. No. 1.2e-81		
Matches 218	Conservative 29	Mismatches 83	Indels 1	Gaps 1
QY	1	MQLVSVLICAVNAEKYFAQSLAAVAGOTWRNLDLIYDGSSTGTPAIRHFOEDDGR	60	
DB	1	MQLVSVLICAVNAEKYFAQSLAAVAGOTWRNLDLIYDGSSTGTPAIRHFOEDDGR	60	
QY	61	RIISPNRLGFIASINIGLDELAKS-GGGEYIARTDADDIASPGWIEKIVGEMEKDRSI	119	
DB	61	RIIAPRNBSGLIPLSINIGLDELAKSGGGEYIARTDADDIAIPMDIEKIVGEMEKDRSI	120	
QY	120	AMGAWLEVLSEENKSVLAATARNCAIMDKPRRHEDIYAVPPFGPIHNNIMKRSYID	179	
DB	121	AMGAWLEVLSEEKDGNRLARHHEHGKIWKPPTRHEDIADFPPGPIHNNIMKRSYID	180	
QY	180	GGLEFDPAYIAEDYKFEYEAAGKGLRLAYPEALVKKYFHHODQSSKLNLOORTARWIK	239	
DB	181	GGLEFARNTERDMWEDYQFWYDYSKGLRLAYPEALVKKYFHHANOVSSKYSIRHETIAOGIO	240	
QY	240	EIRAGYKKAAGIAGVADCLNGLKSTAVLYEKALSGODIGCLRFLVEYFLSELEKYS	299	
DB	241	KIARNDFLOSMGEKTRFDSLEYRQIKAVAYELLEKHILPEDEIARRFLYOCFKRTDLP	300	
QY	300	LTDLDLFTDRVMKRLFAAPQYRKILKMLR	330	
DB	301	AGAWLDFPADGRMRRLFTLRQYFGILHRLK	331	

RESULT 2

RESULT 2

D81027
lacto-N-neotetraose biosynthesis glycosyl transferase Igta NMB1929 [imported] - Neisseria
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: DB1027
R:Retelien, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickley, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzo, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755; PMID:10710307
A:Accession: DB1027
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEXT>
A:Cross-references: GB:AE002541; GB:AE002098; NID:g7227175; PIDN:AAF42258.1; PID:g722718
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1929
C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 60.6%; Score 1069.5; DB 2; Length 349;
Best Local Similarity 65.6%; Pred. No. 3.2e-81;
Matches 217; Conservative 30; Mismatches 83; Indels 1; Gaps 1;

Qy 1 MOPLVSLICAYNAEKYFAQSLAAVVGOTWRNLDILVDDGSTGTPAIARHFOEDGR 60
Db 17 LQPLVSLICAYNAEKYFAQSLAAVVGOTWRNLDILVDDGSTGTPAIARHFOEDGR 76
Qy 61 RIISPRNLGFASINIGDELAKS-GGGEYIARTDADDIASPGWIEKIVGEMEKDRSTI 119
Db 77 RLADPRNSGLPSLINSIGDELAKSGGGEYIARTDADDIAAPDWIEKIVGEMEKDRSTI 136
Qy 120 AMGALEVLSEENKSVLAALARNCAIMDKPTRHEDIYAVEFPGNINHTMIMRSVID 179
Db 137 AMGALEVLSEEKDGNRLARHHEHGIKMKPTRHEDIADFFPGNPINHTMIMRSVID 196
Qy 180 GGLRPDAVIAHEDYKFWYKAGLGLAYVPEALVYKRPFHODTSSKYNLQOORTAWKIK 239
Db 197 GGLRNTEDMAEDYQFWYDVSKLGLRAYVPEALVYKRLHANOVSYSIRQHEIAQGIQ 256
Qy 240 BEIRAGYKKAAGINAGCINLGLKSTAYALYKALSGODIGCLRFLYEFYFSLSEKYS 299
Db 257 KTAARDFLOSMGFRTFDSLEYRQIKAVAYELLEKHLPEDEPARARFLYQCFKRTDPLP 316
Qy 300 LTDLDFLTRVWRKLFAPAQYRKILKKMLR 330
Db 317 AGAWLDFPADGMRRLFTLRQYFGILHRLK 347

RESULT 3
H81970
lacto-N-neotetraose biosynthesis glycosyl transferase NMA0524 [imported] - Neisseria men
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: H81970
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jørgensen, K.; Leather, S.; Mouton, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: AB1775; MUID:20222556; PMID:10761919
A:Accession: H81970
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-346 <PAR>
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83816.1; PID:g737926
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: Igta; NMA0524
C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 58.9%; Score 1039; DB 2; Length 346;

Best Local Similarity 63.3%; Pred. No. 1.1e-78;
Matches 209; Conservative 35; Mismatches 86; Indels 0; Gaps 0;

Qy 1 MOPLVSLICAYNAEKYFAQSLAAVVGOTWRNLDILVDDGSTGTPAIARHFOEDGR 60
Db 1 MOPLVSLICAYNAEKYFAQSLATVAVVOTWRNLELILVDDGSTGTLAIADFOGRDRI 60
Qy 61 RIISPRNLGFASINIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSTI 120
Db 61 KILAQNSGLPSLINSIGDELAKSGGGEYIARTDADDIAAPDWIEKIVGEMEKDRSTI 120
Qy 121 MGAWLEVLSEENKSVLAALARNCAIMDKPTRHEDIYAVEFPGNINHTMIMRSVID 180
Db 121 MGAWLEVLSEEKDGNRLARHHEHGIKMKPTRHEDIADFFPGNPINHTMIMRSVID 180
Qy 181 GGLRPDAVIAHEDYKFWYKAGLGLAYVPEALVYKRPFHODTSSKYNLQOORTAWKIK 240
Db 181 GLRYNTEDMAEDYQFWYDVSKLGLRAYVPEALVYKRLHANOVSYSIRQHEIAQGIQ 240
Qy 241 EIRAGYKKAAGINAGCINLGLKSTAYALYKALSGODIGCLRFLYEFYFSLSEKYS 300
Db 241 TARNDPLOSNGFRTFDSLEYRQIKAVAYELLEKHLPEDEPARARFLYQCFKRTDPPA 300
Qy 301 LTDLDFLTRVWRKLFAPAQYRKILKKMLR 330
Db 301 GAWLDFPADGMRRLFTLRQYFGILHRLK 330

RESULT 4
H64130
glycosyl transferase homolog H11578 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C:Accession: H64130
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: H64130
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-323 <TTGR>
A:Cross-references: GB:U02832; GB:L42023; NID:g1574421; PIDN:AAC23227.1; PID:g1574422
C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 38.7%; Score 682.5; DB 1; Length 323;
Best Local Similarity 45.9%; Pred. No. 4.3e-45;
Matches 151; Conservative 54; Mismatches 107; Indels 17; Gaps 6;

Qy 3 PLVSVLICAYNAEKYFAQSLAAVVGOTWRNLDILVDDGSTGTPAIARHFOEDGR 62
Db 5 PLVSVIYCAVNAEYIDISISIIQTYENLEIIVINGSDTLTSHLEISKIDLRKI 64
Qy 63 ISNPNLGFASINIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSTI 122
Db 65 ISNKNYLFINSINIGLGCFS---GKYFARMDADDIKAPSWIEKIVYVLEKNDHITAMG 120
Qy 123 AMLLEVLSEENKSVLAALARNCAIMDKPTRHEDIYAVEFPGNINHTMIMRSVI -GG 181
Db 121 STLEIIVK-EGGIIISGYKTGDIWKNPDLNHDICEAMLPNHNMTIMKANYRHRK 179
Qy 182 LRFDPVIAHEDYKFWYKAGLGLRAYVPEALVYKRPFHODTSSKYNLQOORTAWKIK 241
Db 180 LIFNNDYVIAEDYKFWYKAGLGLAYVPEALVYKRLHGNQTSVYNNHONETAKIKRE 239
Qy 242 IRAGYKKAAGINAGCINLGLKSTAYALYKALSGODIGCLRFLYEFYFSLSEKYS 301
Db 240 NITVYLNKIGIDIKV-----INSVLEIYHVDKSNK---VLKSTIYEMVMSLDKTYIT 290
Qy 302 DLDFLTRVWRKLFAPAQYRKILKKMLR 330

Db 291 SLHFI--KYHLEFDLKNKNIKKRFR 317

RESULT 5

G95948

Probable glycosyltransferase protein SMB21189 [imported] - Sinorhizobium meliloti (strain C:Species: Sinorhizobium meliloti)

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: G95948

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernat Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: G95948

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-367 <KUR> PIDN:CAC49255.1; PID:915140741; GSPDB:GN00167

A:Cross-references: GB:AL591985; PIDN:CAC49255.1; PID:915140741; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymb

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Laure, hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMB21189

A:Genome: plasmid

Query Match 19.6%; Score 346; DB 2; Length 367;
Best Local Similarity 30.7%; Pred. No. 4.9e-21;
Matches 112; Conservative 53; Mismatches 130; Indels 70; Gaps 13;

3 PLVSLICAYNAEKYFAOSLAAVVGOTWRNLDILIVDGSSTGTPAIARHFOEDGRIRI 62

4 PVSLIVPYNNAEPYIAAIESVLRODYERLEIVAIIDGSDTRSDILERYKSDSRVSI 63

63 ISNPNLGFASLNLGDELAKSGGGEYIARTDADIASPGWIEKIVGEMKRSIIAMG 122

64 ISR-ENRGIATLNGL-ALAK---GELIARMDADDIAPSRSLSQVALFSAEPRLALSG 118

123 AMLEVLSEENKNSVLAALARNCAIMDKPTRHEDIYAVFPFGNPH----- 167

119 TGIDML-----IGNRIINGKP-----NPITRPGSLRILISMEFTIF 153

168 -NNTIMIRRSVI-DGGLRFPAYIHAEYKFEWEAGKLGRLAYPEALVYKRFHODQTS 225

154 MHSIVYNNRVNIPPEMLRNDRPNVHADEPDLFRRITADRPPVIMDEALVAVYIHDSVTS 213

226 KYNLQORTAMKIKEIRI-----GYKAAAGTAVGADCLNTGLKSTAYALYEKALS 277

214 HKRQMRRTHLIVVENIARDAIRLDSALAEIGAIVTSFV--ARLADLVIAL-EREIS 270

278 GO-----DIG--CLRLFEYFLSLEKYSITDLDLDFIDRVRMKFEAFQYKRIKK 327

271 AOPGEVRRAYEDGALCFYFLYQLIAEBQPLTH--EFLITGTGKWGLIRRRERYGLAA 328

328 MLRPV 332

329 ARAPW 333

RESULT 6

B97168

glycosyltransferase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: B97168

R:Noilling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97168

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-333 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80133.1; PID:915025169; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2175

Query Match 16.0%; Score 282.5; DB 2; Length 333;
Best Local Similarity 27.4%; Pred. No. 8.1e-16;
Matches 80; Conservative 59; Mismatches 102; Indels 51; Gaps 10;

4 LVSVLICAYNAEKYFAOSLAAVVGOTWRNLDILIVDGSSTGTPAIARHFOEDGRIRI 63

1 MISVIMPYNCKEYLEESIESILKOTYRDFEFLIYNDGSDNRKSIDINKYANDNRIVVY 60

64 SNPNLGFASLNLGDELAKSGGGEYIARTDADIASPGWIEKIVGEMKRSIIAMGA 123

61 SDDNNKGMVYSLNEGIDR-AK---GSYVARMADADIALPERFERQIEYLKKNKRDVILAC 116

124 WLEVLSEENKNSVLA-----IARNGAIDMKPTRHEDIYAVFPFGNPIHNTMIRRSVI 178

117 KYEAGDVSRREKLEREHYVNDLNS-----ESIESLLEKCYIAHPSVWKMMSVL 168

179 D--GGLRFPAYIHAEYKFEWEAGKLG-RLAYPEALVYKRFHODQ-----SSKY 227

169 KALGVYNLN--YKRTEDYVNLMLRAIAKGYKIAMLEKIRLHNDKTHRDAGFSIR 226

228 NLOQRTAMKIKEIRIAGYKAAAGTAVGADCLNTGLKSTAYALYEKALSCQ 279

227 DIIOQRLEY-VKEKLK-----LDQSFYVIGASNGK 257

RESULT 7

glycosyltransferase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: A97168

R:Noilling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: A97168

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-336 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80132.1; PID:915025168; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2174

Query Match 14.9%; Score 263.5; DB 2; Length 336;
Best Local Similarity 28.2%; Pred. No. 3.1e-14;
Matches 87; Conservative 61; Mismatches 123; Indels 37; Gaps 13;

3 PLVSLICAYNAEKYFAOSLAAVVGOTWRNLDILIVDGSSTGTPAIARHFOEDGRIRI 62

5 PVSLIVPYNNAEPYIAAIESITLDOTYWFDFITVDGSDSTESITNIISYANKDKRIIV 64

63 ISNPNLGFASLNLGDELAKSGGGEYIARTDADIASPGWIEKIVGEMKRSIIAMG 122

65 ISR-EHRGIVDSLNGIN-IAR---GKIYARMDADDISINNIEKQEFLELKNKVDILG 119

123 AMLEVL--EENKNSVLAALARNCAIMDKPTRHEDIYAVFPFGNPI-HNNTMIRRSYD 179

120 TRIEAFGDIDEKQKITV-----NSAFSIR-FDSQNIQEVFLTSCAIIHPSVWFKKDSIV 173

[illegible]

RESULT 8
T44330
glycosyl transferase homolog [imported] - Vibrio cholerae
C:Species: Vibrio cholerae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44330
R:Ryamasaki, S.; Shimizu, T.; Hoshino, K.; Ho, S.T.; Shimada, T.; Nair, G.B.; Takeda, Y.
Gene 237, 321-332, 1999
A:Title: The genes responsible for O-antigen synthesis of Vibrio cholerae O139 are close
A:Reference number: Z22749; MUID:99453293; PMID:10521656
A:Accession: T44330
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-340 <YAM>
A:Cross-references: EMBL:AB012957; NID:g4115688; PIDN:BAA33634.1; PID:g3721684
A:Experimental source: strain 022
C:Genetics:
A:Note: wblc
C:Superfamily: Neisseria meningitidis glycosyl transferase A

[illegible]

RESULT 9
H97167
glycosyltransferase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H97167
R:Noilling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum* ATCC824
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97167
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-33 <KUD>
A:Cross-references: GB:AE001437; PIDN:AAK80131.1; PID:G15025167; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
A:Genetics:

[illegible]

RESULT 10
 BH4114
 expopolysaccharide biosynthesis BH3714 [imported] - *Bacillus halodurans* (strain C-125)
 C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: BH4114
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a
 A:Reference number: AB3650; MUID:20512582; PMID:11058132
 A:Accession: BH4114
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-571 <STO>
 A:Cross-references: GB:AP001519, GB:BA000004, NID:910176109, PIDN:BA07433.1, GSEDB:G
 C:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH3714

Query Match	14.1%	Score 249;	DB 2;	Length 271;
Best Local Similarity	26.7%	Pred. No. 3.8e-13;		
Matches	70;	Conservative 58;	Mismatches 108;	Indels 26; Gaps 9;
QY	1	MQPLSVLICANAEKRYFAOSLAAYVGOTWNNLDLIYDOSSTGCTPRAIARHPOEDGRI	60	
		: : : : : :		
Db	1	MPPEVTALMSYNDKNYISESTESTLNTGFENFEFLIINDSTGSGELLEYSKKKRI	60	
QY	61	RIISPNRNLGFASINIGDELAKSGGSEYIARTDADDIASPGWIEKRYGEMEDRSIIA	120	
		: : : : : :		
Db	61	RIIHNKNNRGSLYSABEV-SLAKA---PWLRADADDVSFKDRLAYOMDHVKAHSELDI	116	
QY	121	MGAWLEVLSEENKNSVLAIALIRNGAIIWDPRIHREDIVVF--PFGNPIHNNTMIIMRSV	177	
		: : : : : :		
Db	117	LGSIYIIDDDKNGNELEIKV-----PTTHKEIALIWTGCF--IHPVLEFKKDSI	164	
QY	178	IDDGLRFDPYAIHAADYKFWTEA--GKLGRLAYYPREALVKYRFHOD--QTSSKYYNLDQR	232	
		: : : : : :		
Db	165	IKAG-SYDRNLRERDDYDLWFLRCLEAKL-KRENDIKPLLYRSDYDYKKNNFVYQQA	222	
QY	233	RTAMKIKEEIRAGYTKKAAGIAY	254	
		: : : : : :		
Db	223	KMGFAGARRVAKAPYAYIGITV	244	

RESULT 11
H64431
glycosyl transferase (EC 2.4.-.-) - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence,revision 13-Sep-1996 #text,change 10-Oct-1997
C:Accession: H64431
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
J.; Reich, C.I.; Overberg, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
Rison, J.D.; Sadov, P.M.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: H64431
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-290 <BUI>
A:Cross-references: GB:067549; GB:L77117; NID:g1591709; PID:g1592318; TIGR:MJ1057; PID:g1592318
C:Genetics:
A:Map position: FOR9696513-997385
:Keywords: glycosyltransferase

	Query Match	14.1%	Score 248.5;	DB: 2;	Length 290;
	Best Local Similarity	30.3%;	Pred. No. 4.5e-13;		
	Matches	79;	Conservative	47;	Mismatches 106; Indels 29; Gaps 11;
OY	2	QPLVSVLICAVN-AEKYFAQSILAAVVGGTWRMLDILI-VDDSGTDGTPALAHHPQEDODGR	59		
	:	: :: :::	:	:	:
Db	7	KPLVSVMATVYPEPKIKESIESIXNOTKDFXFIILKDLPNNKKAEIIEKYQOKDKR	66		
OY	60	IRIIENPRNLRGFIASLINTGLDELAKSNGGEVFARFDADDIASPQIEKIVGSEMEDRSII	119		
	:	: :: :::	:	:	:
Db	67	IIFINERNRLTGGAARN---KAVAIARGKIYAILDADDIALPKLEOFKMEWNRDID	122		
OY	120	AMGANLEVLSEENNKSVALAIAIRNGAIWDKPTRH--EDIVAVPFGNDPIHNNTMIARRSV	177		
	:	: :: :::	:	:	:
Db	123	LIFSVMYFIDENGNI-IKEF-----KPEKYFEKEIKKF--FKRHLYTHPMSMVSK	171		
OY	178	IDGGREFPDPAVTHADYEFKWYEAGKLGRLAYYPALVKYRF-HDDQSSK-----NLL	229		
	:	: :: :::	:	:	:
Db	172	ILKKLKAYDEKLRSDDYDFMFCINDYKFDLIEEFLKLRIPRNDNYLSRIKKOKLSY	231		
OY	230	QOORTAMKIKEEI--RAGIWK	248		
	:	: :: :::	:	:	:
Db	232	YLTKTHMKNKHKHCNNVYFWK	252		

RESULT 12
A12091
glucosyltransferase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Name: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: A12091
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchihara, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A12091
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-343 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB73987.1; PID:917131380; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
::Gene: all12288

Query Match	13.9%	Score 244.5;	DB 2;	Length 343;
Best Local Similarity	30.2%;	Pred. No. 1.2e-12;		
Matches 78;	Conservative 50;	Mismatches 77;	Indels 53;	Gaps 13

[illegible]

RESULT 13
 B75096
 glycosyl transferase PAB0772 - Pyrococcus abyssi (strain Orsay)
 C.Species: Pyrococcus abyssi
 C.Date: 20-Aug-1999 #sequence-revision 20-Aug-1999 #text-change 20-Jun-2000
 C.Accession: B75096
 R.anonymous: Genoscope
 submitted to the EMBL Data Library, July 1999
 A.Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome s
 A.Reference number: A75001
 A.Accession: B75096
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-298 <KAW>
 A.Cross-references: GB:AY248286; GB:AL096836; NID:g5458366; PIDN:CA850071.1; PID:g5454
 A.Experimental source: strain Orsay
 C.genetics:
 A.Gene: PAB0772
 C.superfamily: Neisseria meningitidis glycosyl transferase A

 Query Match 13.8%, Score 243, DB 2, Length 298;
 Best Local Similarity 26.3%, Pred. No. 1.3e-12;
 Matches 83; Conservative 48; Mismatches 128; Indels 56; Gaps 10;

```

QY 2 OPLVLLICAVNAEKKYFPOSTAAVVGOTMRNDLLIVDGSFDPGRALARHNOEDQDGR 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 RPIVSVIIPTRYRANMLRLRALASVYNOKFKDELLIVDASTDNPVEVEST--EDGR 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 IISERNIGFIASLIGLDELAKSGGGEYIARTDADDIASPGWIEKIVEME--KDRSI 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 YLRLEKNSGGPIARNIGI---KRAKGRFIALLDDDDDEMLPRLEEVQYKFEENLKEGCV 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 IAMGMAMLEVLSENNKKSVIUAARNGALIMDK--PRHHDYIAVFFEGNPINHMIMRRS 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 YVGGEY---VSÖDGRILGKRPLRNRDQIYSHLKENFISPTLLIRRE 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 VIDGILRFDPDAYIHAEDYKFWEYEAQKLGRLAYYPALVYKREPHODOTS---SKYMLQOR 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 163 CFKKAAGLFDPRLSSQDDMMMLRIARYKKFEDYVDEILIAKYYVHGQKISFNKKKYPGRER 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 TAWKIKKEELRACGYKWAACIAGVAGDCLNYGL-----KSAVALYKELASGDIGCLR 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 LIRK----HLDIMWNPKI-USIHLSÖMGLLILLSNNTGKGLKTYLTYSIAIAI----- 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 LELYEYFLSEKYSL 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 -----ENLENYMI 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 14

AG1920

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 8.20186 Seconds

(Without alignments)
1704.189 Million cell updates/sec

Title: US-10-007-267-5

Perfect score: 1764

Sequence: 1 MGPLVSLICAYNAEKYFAQ.....APQYRKILKMLRPWKYRSY 337

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	683.5	38.7	323	YF78_HAEIN	Q57287 haemophilus
2	253.5	14.4	290	YAS7_METTA	Q58457 methanococc
3	209	11.8	230	Y868_HAEIN	Q57022 haemophilus
4	195.5	11.1	322	Y586_ANASP	P22639 anabaena sp
5	194	11.0	266	AMSE_ERWAM	Q46635 erwina amy
6	180	10.2	301	AMSB_ERWAM	Q46632 erwina amy
7	178.5	10.1	344	Y1BD_ECOLI	P11290 escherichia
8	167	9.5	446	GGAA_BACSU	P46917 bacillus su
9	164.5	9.3	330	EXOA_RHIME	P33691 rhizobium m
10	160.5	9.1	256	SPSA_BACSU	P33621 bacillus su
11	155	8.8	348	EXOO_RHIME	P33697 rhizobium m
12	150.5	8.5	441	YCDO_ECOLI	P75905 escherichia
13	147	8.3	268	YMDP_BACSU	P36614 bacillus su
14	142	8.0	346	YF20_MYCTU	Q50587 mycobacteri
15	140.5	7.9	267	YG95_HAEIN	Q46215 haemophilus
16	138.5	7.9	299	Y025_MYCPN	P75086 mycoplasma
17	134.5	7.6	900	GGAB_BACSU	P46918 bacillus su
18	131	7.4	342	EXOU_RHIME	P33700 rhizobium m
19	129.5	7.3	909	YAGI_RHISN	P55465 rhizobium s
20	127	7.2	260	DPML_HUMAN	O60702 homo sapien
21	125	7.1	260	DPML_MOUSE	O70152 mus musculu
22	124	7.0	266	DPML_CRIGR	O94983 cricetulus
23	122.5	6.9	299	Y060_MYCPN	P75042 mycoplasma
24	118	6.7	319	EXOW_RHIME	P33702 rhizobium m
25	118	6.7	424	NODC_RHITO	P17862 rhizobium l
26	117.5	6.7	513	YTH1_RHOER	P43770 rhodococcus
27	113.5	6.4	1275	YRBC_MYXXA	O50864 myxococcus
28	112.5	6.4	323	YKCC_BACSU	Q33119 bacillus su
29	112	6.3	294	YG96_HAEIN	Q46214 haemophilus
30	109.5	6.2	279	WCAA_ECOLI	P77414 escherichia
31	109	6.2	236	DPML_SCHPO	O14466 schizosacch
32	109	6.2	319	YF18_MYCTU	Q50590 mycobacteri
33	109	6.2	413	NODC_RHISN	P50357 rhizobium s

34	108.5	6.2	270	1	LPSC_RHIME	Q919m9 rhizobium m
35	107.5	6.1	297	1	Y060_MYCGE	P47306 mycoplasma
36	107.5	6.1	310	1	GTRB_BPP22	P57022 bacterioph
37	107	6.1	241	1	DPML_DROME	O941u7 drosophila
38	107	6.1	434	1	LPGL_LEIDO	O05889 leishmania
39	106	6.0	256	1	YF57_MYCTU	Q50459 mycobacteri
40	106	6.0	331	1	YJ43_SYNY3	P74505 synechocyst
41	105	6.0	182	1	NODC_BRAEL	P53417 bradyrhizob
42	103.5	5.9	243	1	YC22_METUA	Q58619 methanococc
43	103	5.8	612	1	PAG3_CAEEL	P34678 caenorhabdi
44	102.5	5.8	424	1	NODC_RHILY	P04340 rhizobium l
45	101.5	5.8	334	1	ALG5_YEAST	P40350 saccharomyc

ALIGNMENTS

RESULT 1	YF78_HAEIN	STANDARD:	PRT:	323 AA.
ID	YF78_HAEIN			
AC	Q57287: 005077:			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Putative glycosyl transferase H11578 (EC 2.4.1.16).			
GN	H11578.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
OC	Haemophilus.			
OX	NCBI_TaxID=727;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-Rd / KW20 / ATCC 51907;			
RX	MEDLINE-95350630; PubMed-7542800;			
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,			
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,			
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,			
RA	Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,			
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,			
RA	Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,			
RA	Fine L.D., Fitchman J.L., Fuhmann J.L., Geoghegan N.S.M.,			
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,			
RA	Venter J.C.;			
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae			
RL	Rd.";			
CC	Science 269:496-512(1995).			
CC	-I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: U38832; AAC23227.1; -			
DR	TIGR: H11578; -			
DR	InterPro: IPR001173; Glycos_transf_2.			
DR	Pfam: PF00535; Glycos_transf_2; 1.			
KW	Hypothetical protein; Transferase; Glycosyltransferase;			
KW	Complete proteome.			
SO	SEQUENCE 323 AA; 37680 MW; 7C9C2681039A5B4 CRC64;			
Query Match	38.7%; Score 682.5; DB 1; Length 323;			
Best Local Similarity	45.9%; Pred. No. 3.7e+49;			
Matches 151; Conservative 54; Mismatches 107; Indels 17; Gaps 6;				
OY	3 PLVSVLICAYNAEKYFAQSLAVVQGTNRNDILIVDSSTGCTALAHFPODGRRI 62			
DB	5 PLVSVICAYNAEYIDSEISSIINTFTENLITIVINDSGTITVSHLEIKDKRIKI 64			
OY	63 ISNPNLGFIAISLNGIDELAKSGGGEYIARTDADDIASPGWIERKIVGEMERDSIIAMG 122			

```

DB 65 ISKNYMLGFINSLIGLGFCS-----GKYPARMADDDIAKSPWTEIKIVLYLEKNDHTIANG 120
QY 123 ANLEVLISENNKSVLAALRNGAINDKPTRHEDIYAVFPFGPNHNNTMIRRSV-DG6 181
DB 121 SYLEIIVER-ECGIIISQYKRTGIDIMKPNLLHNDICEAMLFYNDPHNNTMIRNRYREHK 179
QY 182 LRPDPAYIAEDYKFWYEGKLGRLAYPEALVKRYFHODQTSKYNLQORTAMKIKEE 241
DB 180 LIFNDYPAEDYKFWSEVSRIGCLANYPEALVKRYLHNGNSSVYNNHONETAKIKRE 239
QY 242 IRAGYKKAAGIANGADCLNLYGLKSTAYALYERKALSGODIGCLRLPYEFYFLEKYSLT 301
DB 240 NITYYLNKIGIDIKY-----INSVLEIHYHDKSNK---VLKSLIYEWYMSLDKXTT 290
QY 302 DLDPLTDVWKRLPAAPQYKRLTKMLR 330
DB 291 SLHFI--KYHLEFLDKONLKIKKFIR 317

```

RESULT 2

```

ID YAS7_METJA STANDARD: PRT: 290 AA.
AC Q58457;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase MJ1057 (EC 2.-.-.-).
GN MJ1057.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Raine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC
CC EMBL: U67549; AAB9061.1; -
CC DR HSSP: P39621; IQGQ.
CC DR TIGR: MJ1057; -
CC DR InterPro: IPR001173; Glycos_transf_2.
CC DR Pfam: PF00535; Glycos_transf_2; 1.
CC KW Hypothetical protein; Transferase; Complete proteome.
CC SEQUENCE 290 AA; 35099 MW; 3f6a1b21c420d74 CRC64;

```

```

Query Match 14.4%; Score 253.5; DB 1; Length 290;
Best Local Similarity 30.7%; Pred. No. 9, 1e-14;
Matches 80; Conservative 51; Mismatches 101; Indels 29; Gaps 12;

```

```

QY 2 QPLVSVLCAVNAEKYPAQSLAAYVGTWRNLID-ILIVDDSGTDPATIAHFQDGR 59
DB 7 KPLVSVVMAATYNEPERKYLKESISIXNQTKDPEFTIIVDNPNNKRAEILIKYQDKR 66

```

```

QY 60 IRISPNRLGFLASLNLIGDELAKSGGEYIARFDADDIASPGWTEIKYGEKEDRSII 119
DB 67 IIFIKNERNLGAGASFNKAVN-IAR---GKIYALDADDIALPKREKPKFMENNRDID 122
QY 120 ANCAMLEVLISENNKSVLAALRNGAINDKPTRH--EDIIVAFPGPNHNNTMIRRSV 177
DB 123 LIFSWMYFIDENGN--ILKEF-----KPEKYKFEIKKYF-FKEHLTVHPSMWVSK 171
QY 178 IDGGLRDPAYIAEDYKRWYEA-GKLGRLAYPEALVYKRF-HODQTSK-----YVL 229
DB 172 ILKKLKYDEKRLSODYDWMRCIANDYKFDILEEFLKRIPNRDNLYSRITKOKLYSY 231
QY 230 OORPAMKIKEE--RAGYWK 248
DB 232 YTLKTKMKKKHFCNNVFWK 252

```

RESULT 3

```

ID Y868_HAEIN STANDARD: PRT: 250 AA.
AC Q57022; P96336;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase H10868 (EC 2.-.-.-).
GN H10868.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC
CC EMBL: U32768; AAC2526.1; -
CC DR HSSP: P39621; IQGQ.
CC DR TIGR: H10868; -
CC DR InterPro: IPR001173; Glycos_transf_2.
CC DR Pfam: PF00535; Glycos_transf_2; 1.
CC KW Hypothetical protein; Transferase; Glycosyltransferase;
CC Complete proteome.
CC SEQUENCE 250 AA; 28915 MW; A5D8220129782B98 CRC64;

```

```

Query Match 11.8%; Score 209; DB 1; Length 250;
Best Local Similarity 22.4%; Pred. No. 3, 6e-10;
Matches 68; Conservative 66; Mismatches 105; Indels 64; Gaps 10;

```

```

QY 3 PLVSVLCAVNAEKYPAQSLAAYVGTWRNLID-ILIVDDSGTDPATIAHFQDGRIRI 62
DB 4 PLISIMPVNAECYLNQILSCINQSYONIEILIDDSGTSKSIIRINNIIDKRVKL 63

```

QY 63 ISNPNLGTASINIGLDELAKSGGEYIARTDADDIASPGWIEKIVGEMEKDRIIANG 122
 DB 64 FLPPTNOGPAARANNIGLER-----AGGDYITFLDSDDFIANDKLEKQNLMLQHLWMTMG 119
 QY 123 AMLEVISENNKSVLAIAIRNGAIWDKPTRHEDIYAVFPFGNDIHNNTIMRRSVIDGCI 182
 DB 120 NFAFCDLEGNQIKLVYT-----SKRIDYTLTLQ-GNQPKIMTYLVERESIK-LI 166
 QY 183 RDPDAVTHADEKFEWYACKLGLRLAYPEALVYRPHQOTSCKYLNQQRRTAMKIKEEL 242
 DB 167 RF-PNIRK-EDYAFELDCIK-----EYKOST 190
 QY 243 RAGYWAAGIAGADCLNYGLKST--AYALYKALSGODICLR-----LELYEFLS 294
 DB 191 LVSHQASSVFRIGKVSNNKFKSAITWTNITRK---EKLGVKSTIYFIYAVNGFTK 247
 QY 295 LEK 297
 DB 248 YRK 250

RESULT 4

YS86_ANASP STANDARD; PRT; 322 AA.
 AC P22639;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative glycosyl transferase alr2836 (EC 2.-.-.-).
 GN ALR2836.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL J. Bacteriol. 172:3131-3137(1990).
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL, AP003591; BAB74535.1; -;
 DR EMBL, AF031959; AAC32401.1; ALT_INIT.
 DR PIR, B35391; B35391.
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2.1.
 KW Hypothetical protein; Transferase; Glycosyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 322 AA; 36388 MW; 753C2FB59327D968 CRC64;

Query Match 11.1%; Score 195.5; DB 1; Length 322;
 Best Local Similarity 23.7%; Pred. No. 6; 5e-09;

Matches 74; Conservative 57; Mismatches 92; Indels 89; Gaps 15;
 QY 5 VSVLICAVNAEKFEAGSLAAVVGQTRNDILIVDGSITDGTGAIRHPEQ-DGRIIRI 63
 DB 3 ISVLIINMYARLSRAINSVLAQTHSDIEITYVDGSDTNSDVTTOLEQAPDKIKPI 62
 QY 64 SNRNLGFTASINIGLDELAKSGGEYIARTDADDIASPGWIEKIVGEMEKDRIIANG 123
 DB 63 FQ-ANOGGAGAFNAG-----AATGEVAFFLDADVDVWPKPHKLORIV-EVFGTSDVGVNH 116
 QY 124 WLEVISENNKSVLAIAIRNGAIWDKPTRHEDIYAV-----PFGNDIHNNTIMR 174
 DB 117 HDII-DGNDKRTIDQASTOG-----PKLSEDLASVILQTNAMCNP-----PTSLAYR 164
 QY 175 RSVIDGGLRDPDAVTHADEYKFW-----YEAGKLERLAYPEALVYRPH 219
 DB 165 REVLKRVFPIDPV-----KRLIWDGCIITCYAFLKIKITLQENLAVYRIRHGANNHMSAA 219
 QY 220 -----OD-----QTSKY-----NLOQRT-----AMKIREIR 243
 DB 220 SATSEQEAQSQAGIENTNOYINDFLRIGGARVDSLRLNQRYRTKYQRSQMDLRE--- 276
 QY 244 AGYKNAAGIAGV 255
 DB 277 --VWGISRLLIG 286

RESULT 5

ANSE_ERWAM STANDARD; PRT; 266 AA.
 ID ANSE_ERWAM
 AC Q46635;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Amylovoran biosynthesis glycosyl transferase anse (EC 2.-.-.-).
 GN ANSE.
 OS Erwinia amylovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OX Erwinia.
 OX NCBI_TaxID=552;
 RN [1]

RP SEQUENCE FROM N.A.
 RX STRAIN=EA1/79;
 RX MEDLINE=95319333; PubMed=7596293;
 RA Bugert P., Geider K.;
 RT "Molecular analysis of the anse operon required for exopolysaccharide
 RT synthesis of Erwinia amylovora.";
 RL Mol. Microbiol. 15:917-933(1995).
 CC -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF AMYLOVORAN WHICH
 CC FUNCTIONS AS A VIRULENCE FACTOR.
 CC
 CC -!- PATHWAY: Exopolysaccharide biosynthesis.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL, X77921; CAA54886.1; -;
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2.1.
 KW Exopolysaccharide synthesis; Transferase; Glycosyltransferase.
 SQ SEQUENCE 266 AA; 30748 MW; 163268A4210EB47B CRC64;

Query Match 11.0%; Score 194; DB 1; Length 266;
 Best Local Similarity 27.0%; Pred. No. 6; 7e-09;
 Matches 80; Conservative 44; Mismatches 128; Indels 44; Gaps 14;

Db 1 MESVLSLTKNEKPEMLEOCLSESLHQOITNADEIVLVDPVSESLKAVATRWANLLPLV 60
 OY 61 RIISNPNRNGFIASLNLGIDELAKSGGEYIARTDADDIASPCWIKRIGYEMEKRSIIA 120
 Db 61 -IVPLEKNIGLGLKALNAGLERCHN---VVARMDTDICLPREFEKQISYMSHEEVL 115
 OY 121 MGAMEVLELSEENKSVLAAIARNGAIWDRPTREDIVAVFPFGNPJHNNTMIR- SYID 179
 Db 116 SCAAV- IERDEHGKERLKR-PLSNNDIHEFARMKNPNHMCVFRKDKVIS 165
 OY 180 GGLRFPDPAITH---AADYKFWYFAGLGR- LAIYPPALVKYRPHQOTSXYLQORRTA 235
 Db 166 AG-----STOHILYMEDYMLRIMSLGHPVANLPVLMKVRAGSDMVN-----KRRG 213
 OY 236 WK-IKEEIRAGYKKAAGIAGADCLNYGLKSPAYALYERKALSGODIGCLRFLYE 290
 Db 214 WNTKSEVOLYRLKLA-----LKOTGFRGTFLYLRMTMLMVKYKMQ-FLYE 261

RESULT 6

AMSB. ERWAM STANDARD; PRT: 301 AA.
 ID AMSB. ERWAM
 AC 046632:
 DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Amylovoran biosynthesis glycosyl transferase amsb (EC 2.-.-.-).
 GN AMSB.
 OS Erwinia amylovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Erwinia.
 OX NCBI_TaxId=552;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-EA1/79:
 RX MEDLINE=95319333; PubMed=7596293;
 RA Bugert P., Geider K.;
 RT "Molecular analysis of the amv operon required for exopolysaccharide
 synthesis of Erwinia amylovora.";
 RT Mol. Microbiol. 15:917-933(1995).
 RL -1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF AMYLOVORAN WHICH
 FUNCTIONS AS A VIRULENCE FACTOR. MAY FUNCTION AS A GLYCOSYL
 TRANSFERASE WHICH TRANSFERS GALACTOSE FROM UDP-GALACTOSE TO A
 LIPID-LINKED AMYLOVORAN-SUBUNIT PRECURSOR.
 CC -1- PATHWAY: Exopolysaccharide biosynthesis.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X77921; CAA54883.1; -
 DR InterPro: IPR001173; Glycos_transf.2.
 DR Pfam: PF00535; Glycos_transf.2.1.
 KM Exopolysaccharide synthesis; Transferase; Glycosyltransferase.
 SQ SEQUENCE 301 AA, 34788 MW, 9900EE080E6EC06 CRC64;

Query Match 10.2%; Score 180; DB 1; Length 301;
 Best Local Similarity 24.4%; Pred. No. 1.1e-07;
 Matches 87; Conservative 51; Mismatches 103; Indels 116; Gaps 21;

OY 6 SVIICAYNAEKTFQASLAIVGQTFRNLDILYDGSSTGTPAIRAHFQDGRIRIISN 65
 Db 7 SVIIPAYNASESITITLDINEOSYKNFVIIVDDKSDA---AQLAE-----VSS 55
 OY 66 PRLNGFIASLNLGIDELAKSGG-----GEYIARTDADDIASPCW----- 104
 Db 56 ERKSGI-KINVLSTKLNGAGARRRGIDLATGTYVCFLDADD---EWHKDKLOQNS 109

OY 105 -IEKIVGEMEK-----DRSIIMAGAMLEVLSEENKSVLAAIARNGAIWDRPTREDIV 157
 Db 110 LIERLEGQDQRRRIIYSQVNIIDSGFLKVMPL-----KRVGHESI 151
 OY 158 AVPEEG-NDIHNNTMIRRSYIDG-LRPDPAYTHAEDYKKEYGKIG-RLAYIPEAL 213
 Db 152 AEYLFQCYGFIQSTIVLKR--DAAEIRFDERYIRHODYLCIRADKLGFKFVMAIQLP 209
 OY 214 VKY-----RF---HQDOTSSKYNLQQRRTAMKIEERACGYWKAAGIAGADCLNYGLKS 266
 Db 210 ANYHWYTRFGSOHKE-SYKYSL-----FW-----LDAMKPHLTR 244
 OY 267 TAYALYKA-----LSGODIGCLRFLYEXFLSLERYSLTDLDPLFDRVYRRL 315
 Db 245 DVTY--KAVKLPRLRYKMDQKSLQA-SLSPARFFFLTKNDNRDPL-----KRLMNL 294

RESULT 7

YIBD_ECOLI STANDARD; PRT: 344 AA.
 ID YIBD_ECOLI
 AC P11290; P22999;
 DT 01-JUL-1989 (rel. 11, Created)
 DT 01-OCT-1994 (rel. 30, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Putative glycosyl transferase yibd (EC 2.-.-.-).
 GN YIBD OR B3615.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxId=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 region from 76.0 to 81.5 minutes.";
 RT Nucleic Acids Res. 22:2576-2586(1994).
 RL [2]
 RP SEQUENCE OF 1-198 FROM N.A.
 RC STRAIN-K12:
 RX MEDLINE=89174812; PubMed=2647748;
 RA Aronson B.D., Somerville R.L., Eperly B.R., Dekker E.E.;
 RT "The primary structure of Escherichia coli L-threonine
 dehydrogenase.";
 RL J. Biol. Chem. 264:5226-5232(1989).
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U00039; AAB18592.1; -
 DR EMBL: AE000439; AAC76639.1; -
 DR EMBL: X06680; CAA29885.1; -
 DR PIR: A30268; Q3ECTH.
 DR PIR: B33276; B33276.
 DR EcoGene: EGI1266; yibd.
 DR InterPro: IPR001173; Glycos_transf.2.
 DR Pfam: PF00535; Glycos_transf.2.1.
 KM Hypothetical protein; Transferase; Glycosyltransferase;
 KM Complete proteome.
 CC CONFLICT 1
 FT 1 MNSTNKLISVIILYNAGDDEFTCAE -> MRAMISALVWK
 FT (IN REF. 2).
 FT CONFLICT 77 77 R -> G (IN REF. 2).
 SQ SEQUENCE 344 AA, 40524 MW, 985DABFE86D5457A CRC64;

Query Match 10.1%; Score 178.5; DB 1; Length 344;
 Best Local Similarity 23.3%; Pred. No. 1.8e-07;

OG Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=94042869; PubMed=8226645;
RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of
RT succinoglycan by Rhizobium meliloti".
RT J. Bacteriol. 175:7033-7044(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=94067019; PubMed=8246891;
RA Becker A., Kleickmann A., Keller M., Arnold W., Puehler A.;
RT "Identification and analysis of the Rhizobium meliloti exoAMONP genes
RT involved in exopolysaccharide biosynthesis and mapping of promoters
RT located on the exoKLAONP fragment".
RT Mol. Gen. Genet. 241:367-379(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhmester J., Chain P.,
RA Vorholter F.J., Hernandez-Lucas I., Becker A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti".
RT Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -I- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF
CC (GLUCOSE). CATALYZES THE FORMATION OF A BETA-1,3 LINKAGE WITH THE
CC GALACTOSE LIPID CARRIER.
CC -I- PATHWAY: Exopolysaccharide biosynthesis.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L20758; AAA16046.1; -
DR EMBL: 222636; CA80345.1; -
DR EMBL: AL603645; CAC49482.1; -
DR PIR: S37027; S37027.
DR PIR: S39956; S39956.
DR InterPro: IPR001173; Glycosyltransf_2.
DR Pfam: PF00535; Glycosyltransf_2.
KW Transferrase; Glycosyltransferase; Exopolysaccharide synthesis;
KW Transmembrane; Plasmid; Complete proteome.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 299 319 POTENTIAL.
SQ SEQUENCE 330 AA; 36167 MW; 6A6A919F152602F1 CRC64;
Query Match 9.3%; Score 164.5; DB 1; Length 330;
Best Local Similarity 24.6%; Pred. No. 2.4e-06;
Matches 67; Conservative 39; Mismatches 105; Indels 61; Gaps 11;

QY 32 NEDLIIVDGSDGTPAIRHQEDGRIRIISPNRLGFIASLNTGLDELAKSGGGEYI 91
DB 39 NARVAVADGSDGTPREIRLATEPRLVFLDNPKRIO--SAAVNRAVAEL--GAGSDYL 95
QY 92 ARTDADDDIASPMIRKIVGEMKDRSIIMGAMLELSE-----NNKSVLAIAIRGATW 147
DB 96 IRIIDAHGIVPDDYCKRLV-----EDALATGADSVVAMQTVGFSTQKATAFAONSKLG 149

QY 148 DKPFRHEDIIVAVFPFGNPIHNNMTIMRRSVIDGGLRFPDPAVIAEDKFWPAGKLG---- 204
DB 150 NGSSKRR--TGAAGHMAEHGHHMLMRLEAKRAVGC--YDSFSHNEDELIDYLRKAGYRI 206
QY 205 -----RLAYYPEALVKYFPHQ-----DQTSSKYNLQOR----- 233
DB 207 WMDTKTSWVYPRAKLVPLFQWYFGYGRGRANFLKRRAMPICRLQMLPLVAPVIAFGALL 266
QY 234 --TAMKKEIRIAGYWKAKGIAGVADCLNYGL 263
DB 267 AIVNM--MAVVPVGMAAA-----CLGIGV 289
RESULT 10
SPSA_BACSU STANDARD; PRT; 256 AA.
AC P39621;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Spore coat polysaccharide biosynthesis protein spsa.
GN SPSA OR IPA-63D;
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaut M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Prescan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees".
RT Mol. Microbiol. 10:371-384(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertello M.G., Bessieres P., Bolotin A., Borchert S.,
RA Brouillet S., Boursier L., Brans A., Braun M., Brignelli S.C., Bron S.,
RA Chou S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Dentzof F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Hajeck J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassatoli A.,
RA Viati A., Wambuit R., Wedler E., Wedler K., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".
RL Nature 390:249-256(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).

CC MEDLINE-99280694; PubMed-10350455;
RT Charnock S.J., Davies G.J.;
RT "Structure of the nucleotide-diphospho-sugar transferase, SpSA from
RT *Bacillus subtilis*, in native and nucleotide-complexed forms.";
RL Biochemistry 38:6380-6385(1999).
CC -!- FUNCTION: GLYCOSYLTRANSFERASE IMPLICATED IN THE SYNTHESIS OF THE
CC SPORE COAT.
CC -!- PATHWAY: Spore coat polysaccharide biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X73124; CAA51619.1; -;
DR EMBL: 299123; CAB15817.1; -;
DR PDB: 10G8; 21-APR-00.
DR PDB: 10G8; 04-MAY-00.
DR PDB: 10G8; 04-MAY-00.
DR Subtilist; Bg10609; spsa.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2.1.
KW Transferase; Glycosyltransferase; 3D-structure; Complete proteome.
FT DISULFID 155 243
SQ SEQUENCE 256 AA; 30184 MW; C20EA9627F5D536B CRC64;

Query Match 9.1%; Score 160.5; DB 1; Length 256;
Best Local Similarity 22.1%; Pred. No. 3.7e-06;
Matches 60; Conservative 61; Mismatches 98; Indels 53; Gaps 12;

QY 3 PLVSVLICAYNAEKYFAOGLAAVVGOTWRNLDILYDSDSTGCTPAIAHFOQDGRIR- 61
DB 2 PLYSVLMTSYNSDYAKSISSLSSTFSDFLIMDDNSNETLVIRPFL-NDNRVRF 60
QY 62 -----IISNRNLGFIASNLINGDELAKSGGEYIARTDADIASPGMIKIVGEM- 113
DB 61 YQSDISGVKEPTEKRYALINQAI-EMAE---GEYITATDNIYMPRLKMWELDTJ 116
QY 114 -KDRSIAMGAELEVISSENNKSVLAIAIRNGA--IWDKPTRHEDIIVAFPPGNPIHNT 170
DB 117 HEKAVIYASASTYHLNE--NRDIYKETVRPAQVYWNAPC-----AIDHCS 161
QY 171 LMRRSVIDG-----GLRDP--PAYIHADYKFWYAGKLGRLAYP-----E 211
DB 162 VMHRSVLEKVEKRGSTYWDSPAFYRIDARFWRVNHF--YPPYPLDELDLNYITDQ 219
QY 212 ALVKYRFHODQTSKYNLQOORTAWRIKEIR 243
DB 220 SIHFQLELEKNEFVRNLPPQNRCLRRESLK 251

RESULT 11
EXOO_RHIME STANDARD; PRT; 348 AA.
ID EXOO_RHIME STANDARD; PRT; 348 AA.
AC P33697;
DT 01-FEB-1994 (Rel. 28. Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Suctionglycan biosynthesis protein exoo (EC 2.---.-).
GN EXOO OR RH1084 OR SMO20959.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=94042869; PubMed=8226645;

RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of
RT suctionglycan by *Rhizobium meliloti*.";
RL J. Bacteriol. 175:7033-7044(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / 5047;
RX MEDLINE=94067019; PubMed=8246891;
RA Becker A., Kleckmann A., Keller M., Arnold W., Puehler A.;
RT "Identification and analysis of the *Rhizobium meliloti* exoAMONP genes
RT involved in exopolysaccharide biosynthesis and mapping of promoters
RT located on the exoKIAMONP fragment.";
RL Mol. Gen. Genet. 241:367-379(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT fixing endosymbiont *Sinorhizobium meliloti*.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -!- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF
CC (GLUCOSE), CATALYZES THE ADDITION OF THE FIFTH SUGAR
CC THE FOURTH AND FIFTH SUGAR.
CC -!- PATHWAY: Exopolysaccharide biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 208.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L20758; AA116044.1; ALT_FRAME.
DR EMBL: 222636; CAA80347.1; -;
DR EMBL: AL603645; CAC49484.1; -;
DR PIR: C49348; C49348.
DR PIR: S37029; S37029.
DR PIR: S39958; S39958.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2.1.
KW Transferase; Glycosyltransferase; Exopolysaccharide synthesis;
KW Plasmid; Complete proteome.
FT DOMAIN 38 45
SQ SEQUENCE 348 AA; 38131 MW; 770B43782F785579 CRC64;

Query Match 8.8%; Score 155; DB 1; Length 348;
Best Local Similarity 25.4%; Pred. No. 1.6e-05;
Matches 60; Conservative 40; Mismatches 104; Indels 32; Gaps 8;

QY 3 PLVSVLICAYNAEKYFAOGLAAVVGOTWRNLDILYDSDSTGCTPAIAHFOQDGRIRI 62
DB 10 PDVTFVAAVYNSADTIVRIEASLAQEGVTEVVVDDCSAATPALVAI--PDPVRVL 67
QY 63 ISNPNLGFIASNLINGDELAKSGGEYIARTDADIASPGWEKIVGEMKDRSIAM- 121
DB 68 IALDRNRGCGGARNAGIG-----AARGRWTAIVLSDDTVPDRIRRRATERADAAGAOIADV 123
QY 122 -----GAMLEVISSENNKSVLAIAIRNGAIWDKPTRHEDIIVAFPPGNPIHNN 169
DB 124 NLDVVSLDGRSLRMFSEAEIARLPQTLPAFTESVLF--RSEHN-----FGYMKRIFE- 175
QY 170 YIMRRSVIDGIRPDPAYIHAEDYKFWYEAGKL-GRLAYPEALVKYRFHODQTS 224
DB 176 ---RRFLENQDLRFDEALRIGEDYILLASALACGRCACVPSAGYIYHIREGSIS 227

```
RESULT 12
YCOG_ECOLI
ID YCOG_ECOLI STANDARD: PRT: 441 AA.
AC P75905;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ycdg.
GN YCOG OR B1022.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97061202; PubMed-8905232;
RA Oshima T., Aliba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitegawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sasaki G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinouchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO Y_PESTIS HMS LOCUS PROTEIN HMSR AND TO
CC S_EPIDERMIDIS ICAM.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-stb.ch).
-----
CC EMBL; AE000204; AAC74107.1; -
CC EMBL; D90739; BAA35803.1; -
CC EMBL; E013663; ycdg.
DR Ecocore: E013663; ycdg.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 351 351 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 393 415 POTENTIAL.
SQ SEQUENCE 441 AA; 50765 MW; 04F5A53D72FEBAB CRC64;

Query Match 8.5%; Score 150.5; DB 1; Length 441;
Best Local Similarity 20.7%; Pred. Iso. 5.1e-05;
Matches 54; Conservative 40; Mismatches 108; Indels 59; Gaps 6;

QY 3 PLVSLICAVNAEKYFAQSIAAVVGOTWRNLILVDDGSTDGTAPALAHFQDGRIRI 62
I : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 75 PSISIIIPCEKNEKEETFIHALAQRYENIEVIAVNDGSTDKTRAILDMAQIPLRLV 134
I : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 ISNPRRLGFIASINIGLDELAKSGGGEYIARTADD-----IASPQWIKIYQEM 113
I : : : : : : : : : : : : : : : : : : : : : : : : : : : :
C
```

```
DB 135 IHEAQNKGAKIALKTG---AAAKSEYLCIDGALLDLDAAAYIVPEMLYNPRGAVT 190
OY 114 KDRSIITAMGAWLEVLSENNKSVLAIAIRNGAIIWMDPTREHDAVVEPFGNPITHNTMIM 173
I : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 191 GNPRIKTRSLVKIYQGEYSIIIGLIKRTQRIYGNVFTYSSVIAAF----- 237
OY 174 RRSVIDGGLRFDPAYIAHEDYKFWYENAGKIGRLAYYPEAL-----VKYRFHODQTSSKY 227
I : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 238 RRSAA-----LAEVGWSDDMITEDIDISMLQLQNMQTIYF 272
OY 228 NLQQRTRAMKIKEEIRAGYWK 248
I : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 273 -EPFALCWIMPEITLKLGMK 291
I : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
YWDF_BACSU
ID YWDF_BACSU STANDARD: PRT: 268 AA.
AC P39614;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase ywdf (EC 2.-.-.-).
GN YWDF OR IPA-56D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE-95020537; PubMed-7934828;
RA Glaser P., Kunst F., Arnaud M., Coudat M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees."
RL Mol. Microbiol. 10:371-384(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE-98044033; PubMed-9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berset M.G., Bessieres P., Bolotin A., Borcherdt S.,
RA Bortis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,
RA Kobayashi Y., Koeltter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivoita C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowski A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Taccoci E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassatelli A.,
RA Viari A., Wambutt R., Wedler H., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumslein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
```

```

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X73124; CA51612.1; -.
CC DR EMBL: 299123; CAB15824.1; -.
CC DR PIR: S39711; S39711.
CC DR Subtilist; B610602; ywdf.
CC DR InterPro: IPR001173; Glycos_transf_2.
CC DR Pfam: PF00535; Glycos_transf_2; 1.
CC DR Hypothetical protein; Transferase; Glycosyltransferase;
CC KW Complete proteome.
CC SEQUENCE 268 AA; 30616 MW; DD8428F7016EC9B3 CRC64;
CC -----
Query Match 8.3%; Score 147; DB 1; Length 268;
Best Local Similarity 25.4%; Pred. No. 5.2e-05;
Matches 65; Conservative 45; Mismatches 120; Indels 26; Gaps 9;
QY 5 VSLICAVNAKRYAQAQSLAAVVGOTWRNLDILYDOSTGTCTPALAHFQODGRIRIIS 64
DB 3 ISIVIVFYNRIPALCELLSISRQILMPYEITIVDAGESEVYPVALY---PELPIAVIN 59
QY 65 NPNRNGFIASINIGDELAKSGGGEYIARTDADIASPGWTEKIVGEKXRSTIAMAM 124
DB 60 LEKNSGVHAARNAAGV---KEASGGCIIMLCDDDFEYTPGHLEKMAKETE-TADVFHSDA- 113
QY 125 LEVLSENNKSVLAIAIRNGAIWDKPTRHEDIVAF---PEGNPINHNMTIMRRSYIDG 180
DB 114 -EIVSEFEKNGRYPVSKR--LFAVTADYED-MRYESTYVYSGS-----MYRFLHDE 162
QY 181 GLRDPATIAHEDKFWYEAQKGLRLAYPPALYKFRHQ--DOTSSKYNNQQRRTANKI 238
DB 163 IGYPDADVHNWMDVFLYRAAKDYKRVKPCASVYAFSDAGDNOSADLGAKRKQYDLRL 222
QY 239 KEIRAGYKWAAGTAV 254
DB 223 SEKHGLGELPTKRNAY 238

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CDP 1551 / Oshkosh;
RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwim M., Half D., Hickey E.,
RA Kolony J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 277826; CAB01396.1; -.
CC DR EMBL: AE007024; AAK45837.1; ALT_INIT.
CC DR TIGR: MT1570; -.
CC DR Tuberculist; RV1520; -.
CC DR InterPro: IPR001173; Glycos_transf_2.
CC DR Pfam: PF00535; Glycos_transf_2; 1.
CC DR Hypothetical protein; Complete proteome.
CC KW SEQUENCE 346 AA; 39578 MW; 22EDA0495E0EFC4 CRC64;
CC -----
Query Match 8.0%; Score 142; DB 1; Length 346;
Best Local Similarity 22.2%; Pred. No. 0.00019;
Matches 60; Conservative 41; Mismatches 109; Indels 60; Gaps 9;
QY 5 VSLICAVNAKRYAQAQSLAAVVGOTWRNLDILYDOSTGTCTPALAHFQODGRIRI 62
DB 1 MSIVISITNQEYIREALDGFQAQRTPEPVEVITLADASTATPRITIGVYARYPQLFRP 60
QY 63 ISNPNLGFISINIGDELAKSGGGEYIARTDADDI-ASPGWTEKIVGEKXRSTIAM 121
DB 61 ILRQTNIGVHANFK---DVLSARGEVLALCEGDDVYTDPLKLSKVQYIDRRPETHVC 116
QY 122 GAMLEVLSENNK-----SVLAIAIRNGAIWDKPTRHEDIVAFPEGNPIN 168
DB 117 FHPVRIYIEDGAKSEPPPLSMWRDLSDVALLARNF-----IQT 155
QY 169 NTMIMRRSVIDIGLRFD--PAYIHAEDYKFWYEAQKGLRLAYPPALYKFRHQDQTSK 226
DB 156 NSVYFRROP-----SYDDIPANVMPIDWYLHVHNAVGGELMALPETMAVYRRH----- 203
QY 227 YNLQORTAMKIKEIRAGYKWAAGTAVGA 256
DB 204 -----AHGIWHSATYTDNRKFEWETRGHGMAA 228

```

```

RESULT 14
YF20 MYCTU STANDARD: PRT: 346 AA.
AC 050587;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein RV1520.
GN RV1520 OR MT1570 OR MTCY195.08C.
OS Mycobacterium tuberculosis.
CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CC NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37Rv;
RX MEDLINE=98295987; Pubmed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).

```

```

RESULT 15
YG95 HAEIN STANDARD: PRT: 267 AA.
AC 048215; 005081;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase H11695 (EC 2.-.-.-).
GN H11695.
OS Haemophilus influenzae.
CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
CC Haemophilus.
CC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A2;
RA McLaughlin R., Abu Kwaik Y., Young R., Spinoia S., Apicella M.;
RT "Characterization and sequence of the lsg locus from Haemophilus
RT influenzae.";

```

RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE-95350630; PubMed-7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McEnaney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd ";
RL Science 269:496-512(1995).
CC -1 SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M94855; AAC24983.1; -;
DR EMBL: U32842; AAC23341.1; -;
DR TIGR: H11693; -;
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
FT CONFLICT 26 26 V -> G (IN REF. 1).
FT CONFLICT 46 46 D -> E (IN REF. 1).
FT CONFLICT 49 49 F -> S (IN REF. 1).
SQ SEQUENCE 267 AA; 30770 MW; A2P1A0532737D8C3 CRC64;

Query Match 8.0%; Score 140.5; DB 1; Length 267;
Best Local Similarity 25.5%; Pred. No. 0.00018;
Matches 63; Conservative 45; Mismatches 114; Indels 25; Gaps 9;

QY 6 SVLICAYNAE--KYFAQSIAAVVQGTWRNLDILYDDGSTDGTPAIARHFOEDGR--IR 61
DB 4 SVLMSLYIKENQFLRECPESLVAQTRQADEIYLVFDGV--TPDIEFYVTEFETKPLK 61
QY 62 IISNPRNIGFIASLNIIGDELAKSGGGEIARTDADDIASPGWIEKIVGEMEKRSIIAM 121
DB 62 LVKLPQNRGLGRALNEGILHC---DYDWVFRMDTDIDICVPRFERQVAFIEQHPSITF 117
QY 122 GAWLEVLSEPNKSVLAIAIRNGAIIWDKPTREHEDIVAVFPFGNPIHNNMTIMRS-VTDG 180
DB 118 GGQIAEFGKNVNDIY-----AYRNVP TSAQETIKFTQKRCPFNHMTVAAYOKSAVIN 169
QY 181 GLRPDPAYIHADYKFWYEAQKLG-RLAYYPBALVKYRFHODOTSRYNLQORRTAMKI- 238
DB 170 GGIED---LQEDYIWMIKLVAGLYMANLPIILYARVNGVSRRCGVNQAKAEMRLF 225
QY 239 KEIRAG 245
DB 226 KLKRLG 232

Search completed: December 2, 2002, 11:56:29
Job time : 10.2019 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 28.9477 Seconds

(without alignments)
2398.732 Million cell updates/sec

Title: US-10-007-267-5

Perfect score: 1764

Sequence: 1 MQLPVSLICAVNAEKYFAQ.....APQYRKILKMLRPWKYRSY 337

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP.ATCHEA:*
2: SP.BACTERIA:*
3: SP.FUNGI:*
4: SP.HUMAN:*
5: SP.INVERTEBRATE:*
6: SP.MAMMAL:*
7: SP.MHC:*
8: SP.ORGANELLE:*
9: SP.PHAGE:*
10: SP.PLANT:*
11: SP.RODENT:*
12: SP.VIRUS:*
13: SP.VERTEBRATE:*
14: SP.UNCLASSIFIED:*
15: SP.VIRUS:*
16: SP.BACTERIAP:*
17: SP.ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1764	100.0	337	2	050949
2	1079	61.2	348	2	09EVD5
3	1074.5	60.9	333	2	051115
4	1069.5	60.6	349	16	09JX06
5	1055	59.8	346	2	093PS0
6	1049	59.5	348	2	050946
7	1048	59.4	346	2	050946
8	1045	59.2	362	2	050951
9	1039	58.9	346	16	09JW66
10	953	54.0	321	2	09L8E9
11	942.5	53.4	322	2	09L8F0
12	712.5	40.4	337	16	09CIR9
13	588	33.3	221	2	093EK6
14	433	24.5	172	2	09RCM9
15	415	23.5	149	2	P96946
16	392	22.2	337	2	034234

17	391	22.2	337	2	P74947	P74947 vibrio chol
18	346	19.6	367	16	092V61	092V61 rhizobium m
19	330.5	18.7	368	16	098JH2	098JH2 rhizobium l
20	282.5	16.0	333	16	097H38	097H38 clostridium
21	275	15.6	117	2	P96944	P96944 neisseria m
22	272	15.4	133	2	093EK8	093EK8 neisseria m
23	265.5	15.1	314	2	09LA88	09LA88 aeromonas h
24	263.5	14.9	336	16	097H39	097H39 clostridium
25	258.5	14.7	340	2	087159	087159 vibrio chol
26	258	14.6	333	16	097H40	097H40 clostridium
27	249	14.1	271	16	09K6L5	09K6L5 bacillus ha
28	244.5	13.9	298	16	08YUP7	08YUP7 anabaena sp
29	243	13.8	298	17	09U216	09U216 pyrococcus
30	237.5	13.5	288	2	056869	056869 yerinia en
31	237	13.4	80	2	09K2R3	09K2R3 neisseria g
32	236	13.4	257	2	093C09	093C09 shigella bo
33	234	13.3	337	16	08YTD6	08YTD6 anabaena sp
34	231	13.1	278	16	P71054	P71054 bacillus su
35	228.5	13.0	298	17	08U2R3	08U2R3 pyrococcus
36	226.5	12.8	301	16	097P78	097P78 streptococ
37	225.5	12.8	318	16	08YSI4	08YSI4 anabaena sp
38	225	12.8	318	16	08YSI3	08YSI3 anabaena sp
39	225	12.8	334	17	058167	058167 pyrococcus
40	224	12.7	732	16	09K6S0	09K6S0 bacillus ha
41	223.5	12.7	311	2	082874	082874 streptococ
42	223	12.6	321	16	08YSM2	08YSM2 anabaena sp
43	223	12.6	333	16	08YWS0	08YWS0 anabaena sp
44	221.5	12.6	324	2	08YLB0	08YLB0 streptococ
45	221	12.5	344	16	P71057	P71057 bacillus su

ALIGNMENTS

RESULT 1	ID	Q50949	PRELIMINARY:	PRT:	337 AA.
AC	Q50949:	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)				
DE	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	Glycosyl transferase.				
GN	LGTD.				
OS	Neisseria gonorrhoeae.				
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
OX	NCBI_TaxID=485;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=F62;				
RX	MEDLINE=95053752; PubMed=7964493;				
RA	Gotschlich E.C.;				
RT	"Genetic locus for the biosynthesis of the variable portion of				
RT	Neisseria gonorrhoeae lipooligosaccharide.";				
RT	J. Exp. Med. 180:2181-2190(1994).				
DR	EMBL: U14554; AAA68012.1;				
DR	Interpro: IPR001173; Glycos. transfr_2.				
DR	Pfam: PF00535; Glycos. transfr_2.1.				
KW	Transferase.				
SQ	SEQUENCE 337 AA: 38393 MW; 5AE21AA75CA732E CRC64;				
Query Match	100.0%; Score 1764; DB 2; Length 337;				
Best Local Similarity	100.0%; Pred. No. 4.8e-137;				
Matches 337; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 MQLPVSLICAVNAEKYFAQSLAAVVGQTRNDLIIYDGSSTDGTPATARFQEDGRT 60				P74947 vibrio chol
DB	1 MQLPVSLICAVNAEKYFAQSLAAVVGQTRNDLIIYDGSSTDGTPATARFQEDGRT 60				092V61 rhizobium m
QY	61 RIISNPRNIGFTASLNTIGDELAKSGGGEYIARTADDDIASGWIKEIGEMEKDRSITA 120				098JH2 rhizobium l
DB	61 RIISNPRNIGFTASLNTIGDELAKSGGGEYIARTADDDIASGWIKEIGEMEKDRSITA 120				097H38 clostridium
QY	121 MCAMLEVLSEENNKSVLAIAIRNGAIWDRPTRHEDIIVAVFPFGNPINHTMTMRSSVIDG 180				P96944 neisseria m
DB	121 MCAMLEVLSEENNKSVLAIAIRNGAIWDRPTRHEDIIVAVFPFGNPINHTMTMRSSVIDG 180				093EK8 neisseria m

```

Db 121 MGAMLEVLSEENKSVLAIAIRNGAIDWKPTRHEDIYAAVFPFGNPJHNTMIMRRSYIDG 180
QY 181 GLRPDPAVYHAEDYKFWYKYGKIGRLAYYPEALVYKRFHODQSSKYNQOORTAKIKE 240
Db 181 GLRPDPAVYHAEDYKFWYKYGKIGRLAYYPEALVYKRFHODQSSKYNQOORTAKIKE 240
QY 241 EIRAGYWKAAAGIAGVADCLNYGLIKSTAYALYKALSGODICGLRFLYEYFLSLEKYSL 300
Db 241 EIRAGYWKAAAGIAGVADCLNYGLIKSTAYALYKALSGODICGLRFLYEYFLSLEKYSL 300
QY 301 TDLLDFLTDVRMKRLFAAPQYRKILKMLRPMKYRSY 337
Db 301 TDLLDFLTDVRMKRLFAAPQYRKILKMLRPMKYRSY 337

RESULT 2
Q9EVD5 PRELIMINARY: PRT; 348 AA.
AC Q9EVD5;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE N-acetylglucosamine transferase.
GN LGTA.
OS Neisseria subflava.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=28449;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21142520; PubMed=11208792;
RA Arking D., Tong Y., Stein D.C.;
RT "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae.";
RL J. Bacteriol. 183:934-941(2001).
DR EMBL: A5240672; AAG09764.1; -.
DR HSB: P39621; I060.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Transferase.
SQ SEQUENCE 348 AA; 40676 MW; BC9E313E9BC0BF41 CRC64;

Query Match 61.2%; Score 1079; DB 2; Length 348;
Best Local Similarity 66.1%; Pred. No. 1.2e-80;
Matches 218; Conservative 29; Mismatches 83; Indels 0; Gaps 0;

```

```

AC Q51115;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Glycosyl transferase.
GN LGTA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58;
RX MEDLINE=96414473; PubMed=8817494;
RA Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;
RT "Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in Neisseria meningitidis.";
RL MOL. Microbiol. 18:729-740(1995).
DR EMBL: U25839; AAC44084.1; -.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
SQ SEQUENCE 333 AA; 38563 MW; 49D8F6CE375387BF CRC64;

Query Match 60.9%; Score 1074.5; DB 2; Length 333;
Best Local Similarity 65.9%; Pred. No. 2.7e-80;
Matches 218; Conservative 29; Mismatches 83; Indels 1; Gaps 1;

```


50 SEQUENCE 346 AA; 40158 MW; 95C89CBA01FEFA9 CRC64;

Query Match 58.9%; Score 1039; DB 16; Length 346;

Best Local Similarity 63.3%; Pred. No. 2.4e-77; Matches 209; Conservative 35; Mismatches 86; Indels 0; Gaps 0;

QY 1 MQLPLSVLICANVKEKFAOSLATAVYVNGTWRNLDILIVDDGSTDGTPTAIAHFOEDGR1 60
DB 1 MQLPLSVLICANVKEKFAOSLATAVYVNGTWRNLDILIVDDGSTDGTPTAIAHFOEDGR1 60
QY 61 RIISPNRLGFIASNLIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRS11A 120
DB 61 RIISPNRLGFIASNLIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRS11A 120
QY 121 MCAWLEVESEENKSVLAIAIRNGA1MDKPTRHEDIVAFEPGPN1HNNTM1RRSV1DG 180
DB 121 MCAWLEVESEENKSVLAIAIRNGA1MDKPTRHEDIVAFEPGPN1HNNTM1RRSV1DG 180
QY 181 GLRPDPAY1HAEDYKFWYEGAKGLRAYVPEALVYRPHODOTSSKYNLQORRTAMK1KE 240
DB 181 GLRYNTERMDWADYQFWYDVSKLGRLAYVPEALVYRLHANVSSKYS1RQHE1AIOG1OX 240
QY 241 EIRAGYWKRAAGIAGVADCLNIGLSTAYALYKALSGODICLRLFLYEYLSLEK1SL 300
DB 241 TARNDFLQSMGKTRPDSLEYROIKAVALLEKHLPEDEFERARFLYQCFKRTDTPPA 300
QY 301 TDLDLDFLTRVVRKLEFAA0YRK1LKMLR 330
DB 301 GAMLDPAADGKMRRLFTMR0YFGL1RLIK 330

RESULT 10
Q9LBE9 PRELIMINARY; PRT; 321 AA.
AC Q9LBE9; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DR 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LGTA.
GN LGTA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS2;
RA Baltnazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT "Mutations in the 1gt operon influence serum-resistance in
RT gonococci.";
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF208059; AAF25877.1; -
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2.1.
SQ SEQUENCE 321 AA; 37052 MW; 6ACADA9A3CB738FA CRC64;

Query Match 54.0%; Score 953; DB 2; Length 321;
Best Local Similarity 62.6%; Pred. No. 2.6e-70;
Matches 191; Conservative 34; Mismatches 80; Indels 0; Gaps 0;

QY 26 VQOTWRNLDILIVDDGSTDGTPTAIAHFOEDGR1RIISPNRLGFIASNLIGLDELAKS 85
DB 1 MNOJWRNLDILIVDDGSTDGTPTAIAHFOEDGR1RIISPNRLGFIASNLIGLDELAKS 60
QY 86 GGGEYIARTDADDIASPGWIEKIVGEMEKDRS11AMGAWLE1LSEKDGKRLAHRH1RG 145
DB 61 GGGEYIARTDADDIASPGWIEKIVGEMEKDRS11AMGAWLE1LSEKDGKRLAHRH1RG 120
QY 146 IMDKPTRHEDIVAFEPGPN1HNNTM1RRSV1DG1RFPDPAY1HAEDYKFWYEGAK1GL 205
DB 121 IMDKPTRHEDIVAFEPGPN1HNNTM1RRSV1DG1RFPDPAY1HAEDYKFWYEGAK1GL 180
QY 206 LAYYPEALVYKRYLHANVSSKHSVRQHE1AIOG1QKTARNDFLQSMGKTRPDSLEYROT 265
DB 301 LYRLIK 306

DB 181 LAYYPEALVYKRYLHANVSSKHSVRQHE1AIOG1QKTARNDFLQSMGKTRPDSLEYROT 240
QY 266 STAYALYKALSGODICLRLFLYEYLSLEK1SLDLDLFDTRVVRKLEFAA0YRK1L 325
DB 241 AAAYELPEKDLPEDEFERARFLYQCFKRTDTPPSGAWLDFAADGRMRLFTLR0YFGL 300
QY 326 KMLR 330
DB 301 YRLIK 305

RESULT 11
Q9LBE9 PRELIMINARY; PRT; 322 AA.
AC Q9LBE9; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DR 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LGTA.
GN LGTA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA899;
RA Baltnazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT "Mutations in the 1gt operon influence serum-resistance in
RT gonococci.";
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF208058; AAF25876.1; -
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2.1.
SQ SEQUENCE 322 AA; 37109 MW; 88FECA5AD41DC2B7 CRC64;

Query Match 53.4%; Score 942.5; DB 2; Length 322;
Best Local Similarity 62.4%; Pred. No. 1.9e-69;
Matches 191; Conservative 34; Mismatches 80; Indels 1; Gaps 1;

QY 26 VQOTWRNLDILIVDDGSTDGTPTAIAHFOEDGR1RIISPNRLGFIASNLIGLDELAKS 85
DB 1 MNOJWRNLDILIVDDGSTDGTPTAIAHFOEDGR1RIISPNRLGFIASNLIGLDELAKS 60
QY 86 -GGGEYIARTDADDIASPGWIEKIVGEMEKDRS11AMGAWLE1LSEKDGKRLAHRH1RG 144
DB 61 GGGEYIARTDADDIASPGWIEKIVGEMEKDRS11AMGAWLE1LSEKDGKRLAHRH1RG 120
QY 145 A1MDKPTRHEDIVAFEPGPN1HNNTM1RRSV1DG1RFPDPAY1HAEDYKFWYEGAK1GL 204
DB 121 A1MDKPTRHEDIVAFEPGPN1HNNTM1RRSV1DG1RFPDPAY1HAEDYKFWYEGAK1GL 180
QY 205 RLAYYPEALVYKRYLHANVSSKHSVRQHE1AIOG1QKTARNDFLQSMGKTRPDSLEYROT 264
DB 181 RLAYYPEALVYKRYLHANVSSKHSVRQHE1AIOG1QKTARNDFLQSMGKTRPDSLEYROT 240
QY 265 KSTAYALYKALSGODICLRLFLYEYLSLEK1SLDLDLFDTRVVRKLEFAA0YRK1 324
DB 241 KAAAYELPEKDLPEDEFERARFLYQCFKRTDTPPSGAWLDFAADGRMRLFTLR0YFGL 300
QY 325 LKMLR 330
DB 301 LYRLIK 306

RESULT 12
Q9CLR9 PRELIMINARY; PRT; 337 AA.
AC Q9CLR9; 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DR 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical protein PM140.
GN PM140.

OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Paustian T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida PM70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AE006155; AAK03224.1;
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2.1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 337 AA; 39267 MW; 8639BCFB5F700DB4 CRC64;

Query Match 40.4%; Score 712.5; DB 16; Length 337;
 Best Local Similarity 46.2%; Pred. No. 1.7e-50;
 Matches 156; Conservative 56; Mismatches 115; Indels 11; Gaps 5;

OY 2 OPLVSLICAVNAEKYEAFOSLAAVVGQTRNDILIVDDGSTGTPAIAHFEODGRIR 61
 DB 6 OPLVSLICAVNAADKIEECIDAIIMOTYKNLEIVVNGSTDTTSLKHYFAGKDPRIK 65
 OY 62 IISPNRLGFIASLNLGDELAKSGGEXIARTDADDIASPGWIEKIVGEMEKDRSIAM 121
 DB 66 IINNENKGFIFASLNLGSIASLNGDYLRARDADDTIKPEWIEKILGYSMLHPQIIAM 121
 OY 122 GAMLEVLSEBNKSVLAIAIRNGATMDKPTREDDIYAVPEPGNPINNTYIMRSYI-DC 180
 DB 122 GSYLTLISDGNLSNLANYHGDENRNPISHREIVEAMLFNPNHNSMIVSTVEREH 181
 OY 181 GUREDPAYTHAEDKKFWEAGKGLRAYPEALVKYRFHODOTSSYNLQOQRTAKIKE 240
 DB 182 GLRFDAVYHTEDTQFWLEYSRIGELANPESLVYRLNHTQTSLSHNKYONIMAKIRK 241
 OY 241 EIRAGYAKAGI--AVGADCLNGLKSTAYALYERAKSGODIGCLRFLEYEFLSEKY 298
 DB 242 RAINVYLQDGLVYHRLGEDIFPHDIEFIOAEL--ASLSLNDGCIIRKILDYCYLSLVDM 298
 OY 299 SLTDLDLFDTRVMKLFPAAPQYRKILKMLRPWKYRS 336
 DB 299 KLINILYFLRDK-NNSYFNKKOKIKIKRIIRPKYES 335

RESULT 13
 OY93EK6 PRELIMINARY; PRT; 221 AA.
 AC O93EK6;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE LGTA.
 GN LGTA.
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=487;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M978;
 RX MEDLINE=21467954; PubMed=11583844;
 RA Zhu P., Klutch M.J., Tsai C.-M.;
 RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
 Expression in Two L8-Immunotype Strains of Neisseria meningitidis";
 RL FEMS Microbiol. Lett. 203:173-177(2001).
 DR EMBL: AF355193; AAL12840.1;
 SQ SEQUENCE 221 AA; 26478 MW; D1P97099B1F6D55F CRC64;

Query Match 33.3%; Score 588; DB 2; Length 221;
 Best Local Similarity 55.3%; Pred. No. 1.7e-40;
 Matches 121; Conservative 25; Mismatches 73; Indels 0; Gaps 0;

OY 112 MEKRSIIMGAMLEVLSEBNKSVLAIAIRNGATMDKPTREDDIYAVPEPGNPINNTM 171
 DB 1 MEKRSIIMGAMLEVLSEBNKSVLAIAIRNGATMDKPTREDDIYAVPEPGNPINNTM 60
 OY 172 IMRSYIDGGLPDPAYTHAEDKKFWEAGKGLRAYPEALVKYRFHODOTSSYNLQO 231
 DB 61 IMRSYIDGGLPDPAYTHAEDKKFWEAGKGLRAYPEALVKYRFHODOTSSYNLQO 120
 OY 232 RSTAMKKEIRAGYAKAGIIVGADCLNGLKSTAYALYERAKSGODIGCLRFLEYEY 291
 DB 121 HELAGSIOGTANNDPFGSNGFTRDSDLEIRQIKAVAYELKHLPEDEFEARARFLVOC 180
 OY 292 FLSLEKYSTLDLDFLTDVVMKLFPAAPQYRKILKMLR 330
 DB 181 FKRTDTLPAGAWLDFEADGRMRRLFTLRQYFGILHRLK 219

RESULT 14
 OY9RG9 PRELIMINARY; PRT; 172 AA.
 AC O9RG9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Glycosyltransferase.
 GN LGTD.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=485;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1291;
 RX MEDLINE=20305049; PubMed=10844691;
 RA Harey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
 RA Phillips N.J., Apicella M.A., Baile M.S.;
 RT "Genococcal lipooligosaccharide is a ligand for the asialoglycoprotein
 receptor on human sperm.";
 RL Mol. Microbiol. 36:1059-1070(2000).
 DR EMBL: AF121135; AAF14362.1;
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2.1.
 KW Transferase.
 SQ SEQUENCE 172 AA; 18872 MW; 7D2737434EDEFCC4 CRC64;

Query Match 24.5%; Score 433; DB 2; Length 172;
 Best Local Similarity 96.6%; Pred. No. 6.6e-28;
 Matches 85; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 MOPVSLICAVNAEKYEFQOSLAAVVGQTRNDILIVDDGSTGTPAIAHFEODGRIR 60
 DB 1 MOPVSLICAVNAEKYEFQOSLAAVVGQTRNDILIVDDGSTGTPAIAHFEODGRIR 60
 OY 61 RITSNPNLGFIFASLNLGDELAKSGG 88
 DB 61 RITSNPNLGFIFASLNLGDELAKSGG 88

RESULT 15
 OY96946 PRELIMINARY; PRT; 149 AA.
 AC P96946;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Glycosyl transferase.
 GN LGTD.
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=487;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=126E;
 RX MEDLINE=20055626; PubMed=10589709;

RA Jennings M.P., Strikanta Y.N., Moxon E.R., Kramer M., Poolman J.T.,
 RA Kuipers B., van der Ley P.;
 RT "The genetic basis of the phase variation repertoire of
 RT lipopolysaccharide immunotypes in *Neisseria meningitidis*.";
 RL Microbiology 145:3013-3021(1999).
 DR EMBL; U65788; AAB48386.1; -.
 DR HSSP; P39621; 10GQ.
 DR InterPro; IPR001173; Glycos_transf_2.
 DR Pfam; PF00535; Glycos_transf_2; 1.
 KW Transferase.
 SQ SEQUENCE 149 AA; 16667 MW; 68A52EB7DF6B552A CRC64;

Query Match 23.5%; Score 415; DB 2; Length 149;
 Best Local Similarity 83.3%; Pred. No. 1.6e-26;
 Matches 85; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

QY 1 MQLPLSVLICAYNAEKYFAQSIAAVVGQTWRLDILIVDDGSTDGPATARHFOEDGRI 60
 DB 1 MQLPLSVLICAYNAEKYFAQSIAAVVGQTWRLDILIVDDGSTDGPATARHFOEDGRI 60
 QY 61 RIISNPNLGFASINIGDELAKSGGGEYIARTDADDIASP 102
 DB 61 RIISNPNLGFASINIGDELAKSGGGEYIARTDADDIASP 102

Search completed: December 2, 2002, 12:02:18
 Job time : 29.9477 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 29.0623 Seconds
(without alignments)
1283.801 Million cell updates/sec

Title: US-10-007-267-6
Perfect score: 1440
Sequence: 1 MGNHVISLASAERRAHAD.....ERQRAELEKVGGRVLFK 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: A_Geneseq_101002.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1440	100.0	280	17	AA91315
2	1425	99.0	280	18	AA96580
3	1036.5	72.0	279	17	AA91312
4	1007	69.9	275	20	AA92156
5	233	16.2	50	20	AA96931
6	144	10.0	1128	22	AB86295
7	129.5	9.0	423	23	AB89453
8	129.5	9.0	517	21	AA841790
9	129.5	9.0	517	22	AA93202
10	129.5	9.0	517	22	AA93206

11	129.5	9.0	578	22	AA681257	Human AFP protein
12	129.5	9.0	579	20	AA930812	Human secreted pro
13	129.5	9.0	595	21	AA812121	Hydrophobic domain
14	129.5	9.0	636	23	AB942016	Human ovarian anti
15	127	8.8	521	21	AA841900	Human ORFX ORF1664
16	127	8.8	622	22	AA93427	Human polypeptide,
17	127	8.8	622	23	AA88452	Human membrane or
18	127	8.8	622	23	AB943477	Human secreted pro
19	109	7.6	739	22	AB811417	Human secreted pro
20	108.5	7.5	168	22	AB820950	Novel human diagno
21	106.5	7.4	201	18	AA920604	H. pylori cytoplas
22	104	7.2	273	22	AA860317	Helicobacter pylor
23	103	7.2	279	18	AA921023	H. pylori cytoplas
24	99	6.9	394	18	AA924257	Aquifex aspartate
25	97	6.7	273	22	AA860321	Helicobacter pylor
26	97	6.7	444	23	AA976659	Helicobacter pylor
27	95	6.6	273	20	AA93253	AA93253
28	94	6.5	767	22	AB811534	Novel human diagno
29	94	6.5	4010	22	AB861520	Drosophila melanog
30	91	6.3	389	22	AA882459	Caenorhabditis ele
31	90.5	6.3	991	16	AA880096	Black widow spider
32	90.5	6.3	1214	16	AA880097	Black widow spider
33	88.5	6.1	161	23	AA815943	Worm C38H2-2/141-3
34	87.5	6.1	657	22	AA933784	Staphylococcus aur
35	87.5	6.1	662	21	AA808632	Amino acid sequenc
36	87.5	6.1	662	21	AA936547	Staphylococcus aur
37	87.5	6.1	1129	19	AA977286	Bovine differentia
38	87.5	6.1	1343	22	AA933259	Novel human secret
39	87	6.0	1674	19	AA881169	Human BAZ1-alpha p
40	86	6.0	291	23	AB849964	Listeria monocytog
41	86	6.0	526	22	AB892861	Human protein sequ
42	85.5	5.9	653	23	AB847763	Listeria monocytog
43	85	5.9	439	22	AA930597	Novel human secret
44	85	5.9	753	20	AA937247	Protein involved i
45	85	5.9	762	22	AA935926	Helicobacter pylor

ALIGNMENTS

RESULT 1	
AA91315	AA91315 standard; Protein; 280 AA.
XX	XX
AC	AA91315:
XX	XX
DT	09-JUL-1996 (first entry)
XX	XX
DE	N. gonorrhoeae glycosyltransferase LgtE.
XX	XX
KW	Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
KW	vaccine.
XX	XX
OS	Neisseria gonorrhoeae strain F62.
XX	XX
PN	MO9610086-A1.
XX	XX
PD	04-APR-1996.
XX	XX
PF	25-SEP-1995; 95WO-US12317.
XX	XX
PR	26-SEP-1994; 94US-0312387.
XX	XX
PA	(UTR) UNIV ROCKEFELLER.
XX	XX
PI	Gotschlich EC;
XX	XX
DR	WPI: 1996-200924/20.
XX	XX
DR	N-PSDB: AAT14061.
XX	XX
PT	Nucleic acids encoding glycosyl transferase(s) - used in the
PT	diagnosis of infection with Neisseria and for the biosynthesis of
PT	oligo:saccharide(s)

```

XX  Claim 12; Fig 2f; 81bp; English.
PS
XX
CC  5 glycosyltransferases (AAR91311-15) are products of the lgt locus
CC  (AAT14061) of Neisseria gonorrhoeae strain F62. Glycosyltransferase
CC  LgtE (AAR91315) can be obtained by expression of the lgtE coding
CC  sequence in recombinant host cells. A method for adding Gal
CC  beta1-4 to GlcNAc or Glc comprises contacting a reaction mixture
CC  containing activated Gal to an acceptor moiety comprising a GlcNAc or
CC  Glc residue in the presence of LgtE. Oligosaccharides can be produced
CC  that, when attached to non-toxic lipids, are useful for Neisseria
CC  vaccine prepns. Blood group core oligosaccharides, and mimics of
CC  lacto-N-neotetraose, gangliosides and saccharide portions of
CC  globoglycolipids can also be produced using the enzymes.
XX
SQ  Sequence 280 AA:

Query Match          100.0%; Score 1440; DB 17; Length 280;
Best Local Similarity 100.0%; Pred. No. 6,4e-148;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 MONHYSLSAABERRAHADTFGSRGIPFOFPDAMPSERLBOAAELVPGLSAHPYLSG 60
DB  1 MONHYSLSAABERRAHADTFGSRGIPFOFPDAMPSERLBOAAELVPGLSAHPYLSG 60
QY  61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAQOFLAEDTWLEERFDKDSAFIVR 120
DB  61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAQOFLAEDTWLEERFDKDSAFIVR 120
QY  121 LETMAKYIVRPDKYLVNENRSFPLESEHCCTAGYIISREAMRFLLDRFAVLPPERIKA 180
DB  121 LETMAKYIVRPDKYLVNENRSFPLESEHCCTAGYIISREAMRFLLDRFAVLPPERIKA 180
QY  181 VDIAMFTFFDEKGPVYOVSPALCTQELHYAKFLSONSMGSDLEKDEKREGRRHRSILK 240
DB  181 VDIAMFTFFDEKGPVYOVSPALCTQELHYAKFLSONSMGSDLEKDEKREGRRHRSILK 240
QY  241 VFEDLKRALGKGRKKRMRORQAELKYGRRVILFK 280
DB  241 VFEDLKRALGKGRKKRMRORQAELKYGRRVILFK 280

RESULT 2
AAM06580
ID  AAM06580 standard; Protein; 280 AA.
XX
AC  AAM06580;
XX
DT  21-MAR-1997 (first entry)
XX
DE  Lipo-oligosaccharide gene-encoded protein.
XX
KM  Polyglycosyltransferase; N-acetylglucosaminyl transferase;
KW  N-acetylglucosaminyl transferase; lipo-oligosaccharide.
XX
OS  Neisseria gonorrhoeae ATCC 33084.
XX
PN  WO9640971-A1.
XX
PD  19-DEC-1996.
XX
PF  03-JUN-1996; 96WO-US08323.
XX
PR  07-JUN-1995; 95US-0478140.
XX
PA  (NEOS-) NEOSE TECHNOLOGIES INC.
XX
PI  Buczala SL, Johnson KF, Roth S;
XX
DR  WPI; 1997-052351/05.
XX
PT  N-PSDB; AAT49230.
XX
PT  Transfer of at least 2 saccharide units using

```

```

PT  poly:glycosyl:transferase - isolated from N. gonorrhoeae, catalyses
PT  the addition of both GlcNAc and GalNAc disaccharide(s) units to a
PT  single galactose moiety
XX
PS  Disclosure: Fig 2F-H; 38bp; English.
XX
CC  A lipo-oligosaccharide-encoding gene region (AAT49230) of Neisseria
CC  gonorrhoeae ATCC 33084 includes coding sequences for 5 proteins
CC  (AAM06576-80), one of which (AAM06576) is a polyglycosyltransferase
CC  that catalyses the addition of GlcNAc and GalNAc disaccharides to
CC  a galactose moiety. The function of the other 4 proteins is not
CC  stated in the specification.
XX
SQ  Sequence 280 AA:

Query Match          99.0%; Score 1425; DB 18; Length 280;
Best Local Similarity 99.3%; Pred. No. 2.7e-146;
Matches 278; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1 MONHYSLSAABERRAHADTFGSRGIPFOFPDAMPSERLBOAAELVPGLSAHPYLSG 60
DB  1 MONHYSLSAABERRAHADTFGSRGIPFOFPDAMPSERLBOAAELVPGLSAHPYLSG 60
QY  61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAQOFLAEDTWLEERFDKDSAFIVR 120
DB  61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAQOFLAEDTWLEERFDKDSAFIVR 120
QY  121 LETMAKYIVRPDKYLVNENRSFPLESEHCCTAGYIISREAMRFLLDRFAVLPPERIKA 180
DB  121 LETMAKYIVRPDKYLVNENRSFPLESEHCCTAGYIISREAMRFLLDRFAVLPPERIKA 180
QY  181 VDIAMFTFFDEKGPVYOVSPALCTQELHYAKFLSONSMGSDLEKDEKREGRRHRSILK 240
DB  181 VDIAMFTFFDEKGPVYOVSPALCTQELHYAKFLSONSMGSDLEKDEKREGRRHRSILK 240
QY  241 VFEDLKRALGKGRKKRMRORQAELKYGRRVILFK 280
DB  241 VFEDLKRALGKGRKKRMRORQAELKYGRRVILFK 280

RESULT 3
AAR91312
ID  AAR91312 standard; Protein; 279 AA.
XX
AC  AAR91312;
XX
DT  09-JUL-1996 (first entry)
XX
DE  N. gonorrhoeae glycosyltransferase LgtB.
XX
KM  Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
KW  vaccine.
XX
OS  Neisseria gonorrhoeae strain F62.
XX
PN  WO9610086-A1.
XX
PD  04-APR-1996.
XX
PF  25-SEP-1995; 95WO-US12317.
XX
PR  26-SEP-1994; 94US-0312387.
XX
PA  (UYRO) UNIV ROCKEFELLER.
XX
PI  Gotschlich EC;
XX
DR  WPI; 1996-200924/20.
XX
PT  P-PSDB; AAT14061.
XX
PT  Nucleic acids encoding glycosyl transferase(s) - used in the
PT  diagnosis of infection with Neisseria and for the biosynthesis of
PT  oligo:saccharide(s)

```

```

XX  Claim 9; Fig 2c; 81pp; English.
PS
XX  5 Glycosyltransferases (AAR91311-15) are products of the lgt locus
CC  (AAR914061) of Neisseria gonorrhoeae strain F62. Glycosyltransferase
CC  LgtB (AAR91312) can be obt'd. by expression of the lgtB coding
CC  sequence in recombinant host cells. A method for adding Gal
CC  beta1-4 to GlcNAc or Glc comprises contacting a reaction mixture
CC  or Glc residue in the presence of LgtB. Oligosaccharides can be
CC  produced that, when attached to non-toxic lipids, are useful for
CC  Neisseria vaccine prep'n. Blood group core oligosaccharides, and
CC  mimics of lacto-N-neotetraase, gangliosides and saccharide
CC  portions of glycolipids can also be produced using the enzymes.
XX
SQ  Sequence 279 AA;

Query Match 72.0%; Score 1036.5; DB 17; Length 279;
Best Local Similarity 71.6%; Pred. No. 4.7e-104;
Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;

OY 1 MONHYSLASAERBAHIAADTFGRGIPQFPFDALMPSERLBOAMAEIVPGLSAHPYLSG 60
DB 1 MONHYSLASAERBAHIAATGSGIPQFPFDALMPSERLERAMAEIVPGLSAHPYLSG 60
OY 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLLEGGAEOFLAEDTWLEERFDKDSAFIVR 120
DB 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLLEGGAEOFLAEDTWLQRFDPDSAFIVR 120
OY 121 LETMFAKYIVRPDKYLVNTNRSFPLLESEHCGTAGIISREAMPFLDFAVLPPERIKR 180
DB 121 LETMFMHVLITSPSGVADYCGRAFPPLLESEHCGTAGIISRKAMRFLDFRAVLPPERLHP 180
OY 181 VDLMEFTYFDEKGPVYOVSPALCTOELHYAKFLSONSMGLSDLEKDEQGR----- 233
DB 181 VDLMEFTYFDEKGPVYOVSPALCTOELHYAKFLHDONSALGSLIEHDLRNKQQRDS 240
OY 234 -----RHRSILKVMFDLKRALGKRGREKKRMREROAROLEKRYGRRVLFK 280
DB 241 PANTFKHRLIRALTKISREKRRQROR-----EDLICKIIVPFQ 279

RESULT 4
ID AAY22156 standard; Protein; 275 AA.
AC AAY22156;
DT 08-SEP-1999 (first entry)
DE N. meningitidis Beta-1,4-galactosyltransferase.
XX
XX  Beta-1,4-galactosyltransferase; lgtB; fusion protein; catalytic domain;
KM  glycosyltransferase; accessory enzyme; nucleotide sugar formation;
KM  saccharide donor; oligosaccharide synthesis;
KM  carbohydrate structure development.
XX
XX  Neisseria meningitidis.
OS
XX  WO9931224-A2.
PN
XX  24-JUN-1999.
PD
XX  15-DEC-1998; 98WO-CA01180.
PF
XX  14-DEC-1998; 98US-0211691.
PR  15-DEC-1997; 97US-0069443.
XX
XX  (CANA ) NAT RES COUNCIL CANADA.
PA
XX  Gilbert M, Makarchuk WW, Young NM;
PI
XX  WPI; 1999-395174/33.

```

```

DR N-PSDB; AAX84281.
XX
XX  A new glycosyltransferase fusion protein useful in the enzymatic
PT  synthesis of oligosaccharides
XX
PS Example 2; Fig 2; 63pp; English.
XX
XX  This sequence represents the Neisseria meningitidis Beta-1,4-
CC  galactosyltransferase (also referred to as lgtB). The invention relates
CC  to a nucleic acid encoding a fusion protein that comprises a
CC  glycosyltransferase catalytic domain and a catalytic domain from an
CC  accessory enzyme that is involved in formation of a nucleotide sugar
CC  which is a saccharide donor for a glycosyltransferase reaction. The
CC  fusion protein is useful in the enzymatic synthesis of oligosaccharides.
CC  The fusion proteins are able to catalyze more than one reaction involved
CC  in the enzymatic synthesis. This is useful for the development of
CC  therapeutic agents that have specific carbohydrate structures.
CC  Carbohydrates are involved in recognition elements on the surface of
CC  cells. The fusion protein can be used for the synthesis of both natural
CC  carbohydrates and synthetic derivatives with novel properties. The fusion
CC  polypeptide allows two glycosyltransferase reactions in a single vessel,
CC  provides improved yields of end products. Additionally, cleanup and
CC  disposal of extra solvents and by-products is reduced. The fusion protein
CC  can also use directly different donor analogues and various acceptors
CC  with a terminal galactose residue.
XX
SQ Sequence 275 AA;

Query Match 69.9%; Score 1007; DB 20; Length 275;
Best Local Similarity 74.7%; Pred. No. 7.5e-101;
Matches 201; Conservative 20; Mismatches 44; Indels 4; Gaps 1;

OY 1 MONHYSLASAERBAHIAADTFGRGIPQFPFDALMPSERLBOAMAEIVPGLSAHPYLSG 60
DB 1 MONHYSLASAERBAHIAADTFGRGIPQFPFDALMPSERLBOAMAEIVPGLSAHPYLSG 60
OY 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLLEGGEEXFLAEDAWLQRFDPDTAFIVR 120
DB 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLLEGGEEXFLAEDAWLQRFDPDTAFIVR 120
OY 121 LETMFAKYIVRPDKYLVNTNRSFPLLESEHCGTAGIISREAMPFLDFAVLPPERIKR 180
DB 121 LETMFMHVLITSPSGVADYCGRAFPPLLESEHGWGTAGIISRKAMRFLDFRAVLPPERLHP 180
OY 181 VDLMEFTYFDEKGPVYOVSPALCTOELHYAKFLSONSMGLSDLEK-----REOGRRR 236
DB 181 VDLMEFTYFDEKGPVYOVSPALCTOELHYAKFLHDONSALGSLIEHDLRNKQQRDS 240
OY 237 RSLKYVFDLKRALGKRGREKKRMREROQ 265
DB 241 PANTFKHRLIRALTKISREKRRQRREQ 269

RESULT 5
ID AAW89331 standard; peptide; 50 AA.
AC AAW89331;
DT 26-FEB-1999 (first entry)
DE Neisseria meningitidis lgtE C-terminal peptide.
XX
XX  Neisseria meningitidis; lgtC; lgtB; beta-1,4-galactosyltransferase;
KM  glycosyltransferase; proteolytic enzyme.
OS
XX  Neisseria meningitidis.
PA
XX  WO9854331-A2.
PN
XX  03-DEC-1998.
PD
XX  26-MAY-1998; 98WO-IB00975.
PF

```

```
XX 27-MAY-1997; 97US-0047751.
PR genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
PA (CANA ) NAT RES COUNCIL CANADA.
PI Wakarchuk WM, Young NM;
DR WPI; 1999-035177/03.
XX
XX Expressing high levels of glycosyltransferases - comprises use of
PT either host cells deficient in proteolytic enzymes or modified
PT glycosyltransferase genes deleted in a proteolytic recognition site
XX
XX Example 1; Fig 8; 61pp; English.
PS
CC A method has been developed of expressing a glycosyltransferase in a
CC host cell. The method comprises introducing into the host cell a nucleic
CC acid encoding the glycosyltransferase and incubating the host cell under
CC conditions appropriate for expression of the glycosyltransferase, where
CC the host cell substantially lacks a protease that cleaves polypeptides
CC between two consecutive positively charged amino acid residues. The
CC glycosyltransferase can be used in in vitro production of
CC oligosaccharide structures which are potential therapeutic agents for
CC use in the manipulation of cell-cell recognition events, particularly
CC adhesion of bacteria and viruses to mammalian cells and leukocyte-
CC endothelial cell interaction through selectins in inflammation. The
CC method provides more readily recoverable active glycosyltransferases
CC than prior art methods involving mammalian glycosyltransferases. The
CC present sequence represents a C-terminal peptide from Neisseria
CC meningitidis 19te from the present invention.
XX
XX Sequence 50 AA:
SQ
XX
XX Query Match 16.2%; Score 233; DB 20; Length 50;
XX Best Local Similarity 94.0%; Pred. No. 9.4e-18;
XX Matches 47; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 231 QGRRHRSRLKVPMDLKRALGKFGREKKRMERQROAELEKVGRYLFFK 280
XX | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1 QERRHRSRLKVPMDLKRALGKFGREKKRMERQROAELEKVGRYLFFK 50
DB
XX
XX RESULT 6
XX ABB62795 standard; Protein; 1128 AA.
XX ID ABB62795
XX AC ABB62795;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 15177.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EM;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL06898.
XX
```

```
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Disclosure; SEQ ID NO 15177; 21pp + Sequence Listing; English.
XX
XX PS The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 1128 AA:
XX
XX Query Match 10.0%; Score 144; DB 22; Length 1128;
XX Best Local Similarity 23.0%; Pred. No. 3.8e-06;
XX Matches 71; Conservative 40; Mismatches 106; Indels 92; Gaps 13;
XX
XX QY 5 VISLASAERRAHIAIDTFGSRGIPQFPDAL---MPSERTEQAMAEIVPGLS---AHFY 57
XX :| | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 853 MINLKRPERERKMERLPEIGIEAHFPAVDGKELSTERLEMGVRLPGYEDPYNHRA 912
XX
XX QY 58 LSGVERACFPMASHAVLWQALDEGLPIYAVFEDVYLGEAGAFQALADTWLEEFDDSAF 117
XX :| | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 913 MTMGELGCFLSHYNIWMVWKQKLEVLIEDD----- 945
XX
XX QY 118 IVRLTETFAKVIIVRPPKVLN-----YENRSEPLLESE---HCG---- 152
XX :| | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 946 -IPEFYFNRMAVR---ILNQANNAOYDLIYGRKRLKESSPAPVENDNLVHAQSYW 1001
XX
XX QY 153 TACYIISREAMREFLDRAVLVPPERIKAYD---LMM-----FTYFDKCGMPYQVS 201
XX :| | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 1002 TLGYVISTLQGA--LKLTLAKPLDKLIPVDEFLPMFDRHPNKTWTEAPKRLVAFSAS 1058
XX
XX QY 202 PALC-----TQELHYAKFLSQNSMLGSD-----LEKDRREGRRHRSLSKYNPDLK-- 246
XX :| | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 1059 PLLLPIYYTGESGYISDPEDSQISETSEGEARLKSDEQVFDHEQEFKLINPELKG 1118
XX
XX QY 247 RALGKFGRE 255
XX :| | | | |
XX DB 1119 ESLSKSHQE 1127
XX
XX RESULT 7
XX ABB89453 standard; Protein; 423 AA.
XX ID ABB89453
XX AC ABB89453;
XX DT 24-MAY-2002 (first entry)
XX DE Human polypeptide SEQ ID NO 1829.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX OS Homo sapiens.
XX PN WO200190304-A2.
XX PD 29-NOV-2001.
XX PF 18-MAY-2001; 2001WO-US16450.
XX
```


PR 19-MAY-2000; 2000US-205515P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Birse CE, Rosen CA;
 PI
 XX WPI: 2002-122018/16.
 DR N-PSDB; ABL89862.
 XX
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -
 PS
 XX Claim 11; SEQ ID NO 1829; 2081pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABL89449-ABL90833) and proteins
 CC (ABB89040-ABB9044) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 423 AA:
 Query Match 9.0%; Score 129.5; DB 23; Length 423;
 Best Local Similarity 25.1%; Pred. No. 3.6e-05;
 Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;
 QY 5 VLSLSAARRRAHIADTFSSRGIPROFPPA---LMPSERLDQAMAEVPGLSAHRY--- 57
 DB 155 VLSLARPPRRRRMLASLWMEISGRVDAVGMWLNSSAIRNLGVLDLPGRQ-DPYSGR 213
 QY 58 -LSGVKACFMSHAVLWEGALDEGLPIYIVFEDDVLG---GSAEOFLAEDPWLERFEX 113
 DB 214 TLTKGVCGLSHYSIWEVVARGLARVLVFEEDVAFESNFGRLERLMEDV-EAEKLSW 272
 QY 114 DSAFIVRLTMEAKVIVRPDKVLYNENRSFPLESEHCCTAGY-----IISREAMR 164
 DB 273 DLIYGRKQ-----VNEKEKFAVEGLPGVIVAGSYVTLAVLALRLAKARKILASQPIR 325
 QY 165 FFL--DRAVVL-----PERRKAVDLMMFTYFEDKMGVYOVSPALCTQELHYA 212
 DB 326 RMLPVDEFLLPIMFDQHPNDQYKA-----HFWPRD-LVAFSAQPLAA-PTHYA 371
 RESULT 8
 AAB41790
 ID AAB41790 standard; Protein; 517 AA.
 AC AAB41790;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF1554 polypeptide sequence SEQ ID NO:3108.
 XX
 KW Human; Open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 OS
 XX Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-0508621.
 PF
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 PI
 XX Shinkets RA, Leach M;
 DR WPI: 2000-602362/57.
 DR N-PSDB: AAC73999.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 2331-2332; 5507pp; English.
 XX
 CC AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 517 AA:
 Query Match 9.0%; Score 129.5; DB 21; Length 517;
 Best Local Similarity 25.1%; Pred. No. 4.8e-05;
 Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;
 QY 5 VLSLSAARRRAHIADTFSSRGIPROFPPA---LMPSERLDQAMAEVPGLSAHRY--- 57
 DB 246 VLSLARPPRRRRMLASLWMEISGRVDAVGMWLNSSAIRNLGVLDLPGRQ-DPYSGR 304
 QY 58 -LSGVKACFMSHAVLWEGALDEGLPIYIVFEDDVLG---GSAEOFLAEDPWLERFEX 113
 DB 305 TLTKGVCGLSHYSIWEVVARGLARVLVFEEDVAFESNFGRLERLMEDV-EAEKLSW 363
 QY 114 DSAFIVRLTMEAKVIVRPDKVLYNENRSFPLESEHCCTAGY-----IISREAMR 164

Db 364 DLIIYGRKQ-----VNPEKETAVEGLPGLVAVAGSYWTLAVALRLAGARKILLASQPLR 416
QY 165 FFL--DREAVL-----PPERIKAVDLMFTYFFDKEMPIYQVSPALCQELHYA 212
Db 417 RMLPVDEFLPIMDOHPNEQYKA-----HFWPRD-LVAFSAQPLLA-PTHYA 462

RESULT 9
AAM93202

ID AAM93202 standard; Protein; 517 AA.
AC AAM93202;
XX
XX 06-NOV-2001 (first entry)
DT
XX
DE Human polypeptide, SEQ ID NO: 2591.
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
XX Homo sapiens.
XX
XX EP130094-A2.
XX
XX 05-SEP-2001.
PD
XX 07-JUL-2000; 2000EP-0114089.
PF
XX 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
PA
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX MPI; 2001-524255/58.
DR N-PSDB; AAK94110.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
XX Claim 8; SEQ ID NO 2591; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 517 AA;

Query Match 9.0%; Score 129.5; DB 22; Length 517;
Best Local Similarity 25.1%; Pred. No. 4.8e-05;
Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

QY 5 VISLASAERRAHIAITFGSRGIPFOFFDA----LMPSERLQAMAEIVPGLSAHRY--- 57
Db 246 VISLARRPDRERRLASLWMEISGRVDAVDGMLNNSAIRMLGYDLPLPGYO-DPYSGR 304
QY 58 -LSGVKACFMSHAVLMEQALDEGLPIYAVFEDDVLLG---EGAEQFLAEDPTMLEERFDK 113
Db 305 TLTGKEVGCFLSHYSTWEEVYANGRLARVLVFEEDVAFESNFRGLRLMEDV-EAEKLSW 363
QY 114 DSAFIVRLTETMAKVIVRPDKVLNENRSFPLESEHCHTAGY-----IISREAMR 164

Db 364 DLIIYGRKQ-----VNPEKETAVEGLPGLVAVAGSYWTLAVALRLAGARKILLASQPLR 416
QY 165 FFL--DREAVL-----PPERIKAVDLMFTYFFDKEMPIYQVSPALCQELHYA 212
Db 417 RMLPVDEFLPIMDOHPNEQYKA-----HFWPRD-LVAFSAQPLLA-PTHYA 462

RESULT 10
AAM93206

ID AAM93206 standard; Protein; 517 AA.
AC AAM93206;
XX
XX 06-NOV-2001 (first entry)
DT
XX
DE Human polypeptide, SEQ ID NO: 2599.
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
XX Homo sapiens.
XX
XX EP130094-A2.
XX
XX 05-SEP-2001.
PD
XX 07-JUL-2000; 2000EP-0114089.
PF
XX 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
PA
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX MPI; 2001-524255/58.
DR N-PSDB; AAK94114.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
XX Claim 8; SEQ ID NO 2599; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 517 AA;

Query Match 9.0%; Score 129.5; DB 22; Length 517;
Best Local Similarity 25.1%; Pred. No. 4.8e-05;
Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

QY 5 VISLASAERRAHIAITFGSRGIPFOFFDA----LMPSERLQAMAEIVPGLSAHRY--- 57
Db 246 VISLARRPDRERRLASLWMEISGRVDAVDGMLNNSAIRMLGYDLPLPGYO-DPYSGR 304
QY 58 -LSGVKACFMSHAVLMEQALDEGLPIYAVFEDDVLLG---EGAEQFLAEDPTMLEERFDK 113
Db 305 TLTGKEVGCFLSHYSTWEEVYANGRLARVLVFEEDVAFESNFRGLRLMEDV-EAEKLSW 363
QY 114 DSAFIVRLTETMAKVIVRPDKVLNENRSFPLESEHCHTAGY-----IISREAMR 164

Db 364 DLIIYGRQ-----VPEKETAVEGLPGILVAGSYWTLAYALRLAGARKLASQPLR 416
 QY 165 FFL--DREAVL-----PPERIKAVDLMFTYFDEKGMVYVSPALCTQELHYA 212
 Db 417 RMLPVDEFLLPMEFDQHPNDQYKA-----HFWPRD-LVAFSAQPLLAA-PTHYA 462

RESULT 11

AAG81257
 ID AAG81257 standard; Protein: 578 AA.

AC AAG81257;

DT 10-SEP-2001 (first entry)

XX Human AFP protein sequence SEQ ID NO:32.

XX Human AFP protein sequence SEQ ID NO:32.

XX Human, secreted protein; secretion; bacterial cell; fungal cell;

XX eukaryotic cell; fusion protein; maltose binding protein;

XX immunoglobulin constant region; polynhistidine tag.

OS Homo sapiens.

PN WO200129221-A2.

PD 26-APR-2001.

PE 20-OCT-2000; 2000WO-US29052.

PR 20-OCT-1999; 99US-0160712.

XX (ZYMO) ZYMOGENETICS INC.

PI Conklin DC, Yee DP;

DR WPI: 2001-300340/31.

XX N-PSDB; AAH52108.

XX Isolated polypeptide for directing secretion of proteins of interest

XX from a host cell including, e.g. bacteria, includes contiguous amino

XX acid residues of polypeptide with specified amino acids

XX Claim 1; Page 121-122; 617pp; English.

XX AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242

XX to AAG81453. The secreted proteins can be used for directing the

XX secretion of proteins of interest from a host cell including bacteria,

XX fungal cells, and cultured higher eukaryotic cells. The present invention

XX also describes fusion proteins, where a secreted protein of the invention

XX is operably linked via a peptide bond or peptide linker to a second

XX protein selected from the group consisting of maltose binding protein,

XX an immunoglobulin constant region, a polynhistidine tag and a peptide

XX given in AAG81453.

XX Sequence 578 AA;

Query Match 9.0%; Score 129.5; DB 22; Length 578;
 Best Local Similarity 25.1%; Pred. No. 5.6e-05;
 Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

QY 5 VLSAASAEARRAHIDTFEGSRGIPROFEDA---LMPSERLQAMAEVPGLSAHFY--- 57
 Db 324 VLSLRPDRRRRRMLASLMEISGRVDAVDGWMNLSAINLGVLDLPGVQ-DPYSGR 382
 QY 58 -LSGVKACFMSHNAVMEQALDEGLPIYAFVDDVILG--EGADQFLAEDTWLEERFDK 113
 Db 383 TLTKGVGCEFLSHYSIMEEVAVARGLARVLFEDVAFESNFRGLERLMEDV-EAEKLSM 441
 QY 114 DSAPFLRLTETMAKAVYVRDXYLVNENRSPILLESNHCSTAGY-----IISREMR 164
 Db 442 DLIIYGRQ-----VPEKETAVEGLPGILVAGSYWTLAYALRLAGARKLASQPLR 494

QY 165 FFL--DREAVL-----PPERIKAVDLMFTYFDEKGMVYVSPALCTQELHYA 212
 Db 495 RMLPVDEFLLPMEFDQHPNDQYKA-----HFWPRD-LVAFSAQPLLAA-PTHYA 540

RESULT 12

AAV30812
 ID AAV30812 standard; Protein: 579 AA.

AC AAV30812;

DT 12-OCT-1999 (first entry)

XX Human secreted protein encoded from gene 2.

XX Secreted protein; prevention; treatment; protein therapy; gene therapy;

XX diagnosis; cancer; tumour; neurodegenerative disorder; blood disorder;

XX developmental abnormality; fetal deficiency; leukemia; autoimmune; acne;

XX hepatic disease; renal disease; lymphoma; inflammation; allergy; asthma;

XX Alzheimer's disease; cognitive disorder; schizophrenia; obesity; sepsis;

XX osteoporosis; arthritis; infection; AIDS; connective tissue disorder;

XX transplant rejection; diabetes; psoriasis; cardiovascular disorder;

XX reproductive disorder; food additive; food preservative; human; primer;

XX early promoter; GAS; gamma activation element.

OS Homo sapiens.

PN WO9940100-A1.

PD 12-AUG-1999.

PE 04-FEB-1999; 99WO-US02293.

PR 09-FEB-1998; 98US-0074341.

PR 09-FEB-1998; 98US-0074037.

PR 09-FEB-1998; 98US-0074118.

PR 09-FEB-1998; 98US-0074141.

PR 09-FEB-1998; 98US-0074157.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kyaw H, Lafleur DW, Moore PA, Rosen CA, Ruben SM;

XX Shi Y, Wei Y;

XX WPI: 1999-479426/40.

XX N-PSDB; AAZ00803.

XX New isolated human genes potentially useful for, e.g. developmental

XX abnormalities and fetal deficiencies

XX Claim 1b; Page 206-208; 263pp; English.

XX This invention describes novel isolated human genes and the secreted

XX proteins they encode. The polynucleotides and their corresponding

XX secreted polypeptides are useful for preventing, treating or ameliorating

XX medical conditions e.g. by protein or gene therapy. Also pathological

XX conditions can be diagnosed by determining the amount of the new

XX polypeptides in a sample or by determining the presence of mutations in

PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX
 PS Claim 11: SEQ ID NO 3148; 2922pp: English.
 XX
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABO54131-ABO56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovarian and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_sequences.

XX Sequence 636 AA;

Query Match 9.0%; Score 129.5; DB 23; Length 636;
 Best Local Similarity 25.1%; Pred. No. 6.4e-05;
 Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

OY 5 VISLASAERRRAHIAADFGSGCIPQPFDA---LMSERLEQAMAEIVPGLSAHPY--- 57
 DB 365 VISLARPRDRERREMLASIMEMEISGRVVDADVGMMLNSSAIRNLGVDDLPGYQ-DPYSGR 423
 OY 58 -LSGVEKACFEMSHAVIMEQALDEGLPYIAVEDDYILG---EGAEQFLAEPTWLEERDK 113
 DB 424 TLTKGEVGCFLSHSIMEEVARGLARVLVEDDVRRESNFRGLERLMDV-EAEKLSW 482
 OY 114 DSAFIVRLTFMFAVIVAPDKVLNENRSPFLSEHCGTAGY-----IISREAMR 164
 DB 483 DLITLGRQ-----VNEPEKTAIVEGLPGILVAVAGYSWTIATALRLAGAKKLASQPLR 535
 OY 165 FFL--DREAVL-----PERRIKAVDLMMFTYFDEKGMVQVSPALCTOELHYA 212
 DB 536 RMLPVDELPIIMFDQHNEQYKA-----HFMPRD-LVAFSAQPLLA-PTHYA 581

RESULT 15

AAAB41900 standard; protein: 521 AA.

XX AAAB41900;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF1664 polypeptide sequence SEQ ID NO:3328.

KW Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;
 KW antiviral; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteoprotic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antiplatelet;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

OS Homo sapiens.

PN WO200058473-A2.

XX 05-OCT-2000.

PD 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach M;

XX WPI: 2000-602362/57.

XX N-PSDB: AAC76109.

PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -

PS Claim 11: Page 2519-2520; 5507pp: English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 521 AA;

Query Match 8.8%; Score 127; DB 21; Length 521;
 Best Local Similarity 23.8%; Pred. No. 9e-05;
 Matches 69; Conservative 45; Mismatches 120; Indels 56; Gaps 15;

OY 5 VISLASAERRRAHIAADFGSGCIPQPFDA---MSEERLEQAMAEIVPGLSAHPY--- 57

DB 244 MINLRRODRERREMLRALQAOEIECLVEAVDGRAMNTSQVEALCIQMLPY-RDPYHGR 302

OY 58 -LSGVEKACFEMSHAVIMEQALDEGLPYIAVEDDYILGEGAEQF-----LAETWLEE 109

DB 303 PLTKGELGCLFSLHNINIMKEVVDRLQKSLVEEDDRF---EIFKRLMLNIRV-ERE 357

OY 110 RFDKDSAFIVRLTFMFAVIVAPDKVLNENRSPFLSEHC-GTAGYIISREAMRFLD 168

Db	358	GLDMDLIYVGKRMQ----	VENHREKAY----	PRVRNLIYEADYUWYTLAVYISLGARKLL-	409
Oy	169	RFAYVLPERRIKAVDLMFTFFEDKEGRVYQVSPALSTOELH-----	YAKFLSÖNS	219	
Db	410	--AAEPLESKMLPVDEFL--PVWFEDHN--PVSEYKANFSLNMLHAFSVEPLLIYTPHTHYGDQ	464		
Oy	220	MLGSDLEKDEKROEGRRHRSRLKVMEDLKRALGKFGEREKKRMERORAELE	269		
Db	465	GYVSDTEFSVYVNNH-----YKTDMD-----	RAKSÖKMRQÖALSTRE	502	

```
Search completed: December 2, 2002, 11:59:04
Job time : 31.0623 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 : Search time 10.4223 Seconds
(without alignments)
790.458 Million cell updates/sec

Title: US-10-007-267-6
Perfect score: 1440
Sequence: 1 MONHVISLASAERRAHAD.....ERQROALEKXVGRVILFK 280

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCUTS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	ID	Description
1	1440	100.0	280	1 US-08-312-387B-6	Sequence 6, Appli
2	1440	100.0	280	1 US-08-683-426-6	Sequence 6, Appli
3	1440	100.0	280	1 US-08-683-458-6	Sequence 6, Appli
4	1440	100.0	280	2 US-08-878-360-6	Sequence 6, Appli
5	1440	100.0	280	3 US-08-478-140B-6	Sequence 6, Appli
6	1440	100.0	280	4 US-09-333-412-6	Sequence 6, Appli
7	1440	100.0	280	4 US-09-338-943-6	Sequence 6, Appli
8	1036.5	72.0	279	1 US-08-312-387B-8	Sequence 8, Appli
9	1036.5	72.0	279	1 US-08-683-426-8	Sequence 8, Appli
10	1036.5	72.0	279	1 US-08-683-458-8	Sequence 8, Appli
11	1036.5	72.0	279	2 US-08-878-360-8	Sequence 8, Appli
12	1036.5	72.0	279	4 US-09-333-412-8	Sequence 8, Appli
13	99	6.9	394	2 US-08-646-580B-40	Sequence 40, Appli
14	99	6.9	394	4 US-09-412-184-40	Sequence 40, Appli
15	83	5.8	712	2 US-08-468-576B-17	Sequence 17, Appli
16	83	5.8	712	2 US-08-468-579B-17	Sequence 17, Appli
17	83	5.8	712	3 US-08-468-577B-17	Sequence 17, Appli
18	81.5	5.7	664	1 US-08-485-284A-5	Sequence 5, Appli
19	79.5	5.5	1098	4 US-08-923-982A-8	Sequence 8, Appli
20	78.5	5.5	1104	4 US-08-923-992A-4	Sequence 4, Appli
21	78	5.4	1388	4 US-09-134-001C-3799	Sequence 3799, Ap
22	76.5	5.3	1388	4 US-09-572-191-2	Sequence 2, Appli
23	76.5	5.3	1388	4 US-09-723-262-2	Sequence 2, Appli
24	76.5	5.3	1388	4 US-09-723-219-2	Sequence 2, Appli
25	75.5	5.2	1128	4 US-08-923-992A-6	Sequence 6, Appli
26	75.5	5.2	1227	1 US-08-448-170-8	Sequence 8, Appli
27	75.5	5.2	1227	4 US-08-961-803-9	Sequence 9, Appli

28	75	5.2	337	3 US-09-032-372-2	Sequence 2, Appli
29	75	5.2	1375	4 US-09-722-139-2	Sequence 2, Appli
30	75	5.2	1375	4 US-09-721-832-2	Sequence 2, Appli
31	75	5.2	1375	4 US-09-721-689-2	Sequence 2, Appli
32	74.5	5.2	506	4 US-09-134-001C-4049	Sequence 4049, Ap
33	74	5.1	344	6 5210183-2	Patent No. 5210183
34	74	5.1	630	4 US-09-360-545-67	Sequence 67, Appli
35	74	5.1	683	6 5210183-3	Patent No. 5210183
36	73.5	5.1	487	4 US-09-724-224-8	Sequence 8, Appli
37	73.5	5.1	512	4 US-09-724-224-4	Sequence 4, Appli
38	73.5	5.1	539	3 US-09-057-969-4	Sequence 4, Appli
39	73.5	5.1	624	3 US-09-057-969-3	Sequence 3, Appli
40	73.5	5.1	954	3 US-09-057-969-2	Sequence 2, Appli
41	73.5	5.1	984	1 US-08-242-932-2	Sequence 2, Appli
42	73.5	5.1	984	1 US-08-714-481-2	Sequence 2, Appli
43	73.5	5.1	984	5 PCT-US95-06111-2	Sequence 2, Appli
44	73.5	5.1	1164	4 US-08-923-992A-2	Sequence 2, Appli
45	73	5.1	969	1 US-08-365-689-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-08-312-387B-6
: Sequence 6, Application US/08312387B
: Patent No. 5545553
:
: GENERAL INFORMATION:
: APPLICANT: Gotschlich, Emil C.
: TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
: TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/312,387B
: FILING DATE: July 7, 1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 600-1-095
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 487-5800
: TELEFAX: 201 443-1684
:
: INFORMATION FOR SEQ ID NO.: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 280 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-08-312-387B-6
:
: Query Match 100.0%; Score 1440; DB 1: Length 280;
: Best Local Similarity 100.0%; Pred. No. 7e-155;
: Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 MONHVISLASAERRAHADTGTGRCIPPOFPDAMPSTRLEQNAAEVPGISAPRYSG 60
DB 1 MONHVISLASAERRAHADTGTGRCIPPOFPDAMPSTRLEQNAAEVPGISAPRYSG 60
QY 61 VKACFMASHAVLMEQALDEGLPYIAVFEDDVLGCGAQOFLAEDTWLBERFDKDSAFYR 120

Db 61 VERACFMSHAVIWEQALDGLPIYIAVEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
QY 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180
Db 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180
QY 181 VDLMMFTYFDEKGMFVYQVSPALCTQELHYAKFLSÖNSMLGSDLEKDEÖGRHRRSLK 240
Db 181 VDLMMFTYFDEKGMFVYQVSPALCTQELHYAKFLSÖNSMLGSDLEKDEÖGRHRRSLK 240
QY 241 VMFDLKRALGKFGREKKRMRÖRQAELEKVGGRVILFK 280
Db 241 VMFDLKRALGKFGREKKRMRÖRQAELEKVGGRVILFK 280

RESULT 2
US-08-683-426-6
; Sequence 6, Application US/08683426
; Patent No. 5705367

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,426
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-426-6

Query Match 100.0%; Score 1440; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 7e-155;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MONHVISLSAERRAHINDTFGSRGIPQFPDAMPSERLBOAMAEIVPGLSAHPYLSG 60
Db 1 MONHVISLSAERRAHINDTFGSRGIPQFPDAMPSERLBOAMAEIVPGLSAHPYLSG 60
QY 61 VERACFMSHAVIWEQALDGLPIYIAVEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
Db 61 VERACFMSHAVIWEQALDGLPIYIAVEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

QY 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180
Db 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180
QY 181 VDLMMFTYFDEKGMFVYQVSPALCTQELHYAKFLSÖNSMLGSDLEKDEÖGRHRRSLK 240
Db 181 VDLMMFTYFDEKGMFVYQVSPALCTQELHYAKFLSÖNSMLGSDLEKDEÖGRHRRSLK 240
QY 241 VMFDLKRALGKFGREKKRMRÖRQAELEKVGGRVILFK 280
Db 241 VMFDLKRALGKFGREKKRMRÖRQAELEKVGGRVILFK 280

RESULT 3
US-08-683-458-6
; Sequence 6, Application US/08683458
; Patent No. 5798233

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,458
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REFERENCE/DOCKET NUMBER: 600-1-095A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-458-6

Query Match 100.0%; Score 1440; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 7e-155;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MONHVISLSAERRAHINDTFGSRGIPQFPDAMPSERLBOAMAEIVPGLSAHPYLSG 60
Db 1 MONHVISLSAERRAHINDTFGSRGIPQFPDAMPSERLBOAMAEIVPGLSAHPYLSG 60
QY 61 VERACFMSHAVIWEQALDGLPIYIAVEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
Db 61 VERACFMSHAVIWEQALDGLPIYIAVEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
QY 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180
Db 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180

QY 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSJK 240
| | | | |
DB 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSJK 240

QY 241 VMFDLKRALGKFGREKKRMRERQROALEKYYGRVILFK 280
| | | | |
DB 241 VMFDLKRALGKFGREKKRMRERQROALEKYYGRVILFK 280

RESULT 4

US-08-878-360-6
; Sequence 6, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,360
; FILING DATE: 18-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,426
; FILING DATE:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-0958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5600
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-878-360-6

Query Match 100.0%; Score 1440; DB 2; Length 280;

Best Local Similarity 100.0%; Pred. No. 7e-155;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MONHVISLSAABRRRAHIDTFGSRGIPQFPDAMPSERLLEQAMAEIVPGLSAHPYLSG 60
| | | | |
DB 1 MONHVISLSAABRRRAHIDTFGSRGIPQFPDAMPSERLLEQAMAEIVPGLSAHPYLSG 60

QY 61 VERACFMSHAIVLEQALDEGLPIYIAVEEDVLLGEGAEQFLADDTWLEERFSDSAFIYA 120
| | | | |
DB 61 VERACFMSHAIVLEQALDEGLPIYIAVEEDVLLGEGAEQFLADDTWLEERFSDSAFIYA 120

QY 121 LETMFAKVIYRPDKVYNTENRSPFLLESEHCAGTIIISREARFPLDRFVAVLPERRIKA 180
| | | | |
DB 121 LETMFAKVIYRPDKVYNTENRSPFLLESEHCAGTIIISREARFPLDRFVAVLPERRIKA 180

QY 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSJK 240
| | | | |
DB 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSJK 240

QY 241 VMFDLKRALGKFGREKKRMRERQROALEKYYGRVILFK 280
| | | | |
DB 241 VMFDLKRALGKFGREKKRMRERQROALEKYYGRVILFK 280

RESULT 5

US-08-478-140B-6
; Sequence 6, Application US/08478140B
; Patent No. 6127153
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: ROTH, STEPHEN
; APPLICANT: BUZZALA, STEPHANIE L.
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,140B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-140B-6

Query Match 100.0%; Score 1440; DB 3; Length 280;

Best Local Similarity 100.0%; Pred. No. 7e-155;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MONHVISLSAABRRRAHIDTFGSRGIPQFPDAMPSERLLEQAMAEIVPGLSAHPYLSG 60
| | | | |
DB 1 MONHVISLSAABRRRAHIDTFGSRGIPQFPDAMPSERLLEQAMAEIVPGLSAHPYLSG 60

QY 61 VERACFMSHAIVLEQALDEGLPIYIAVEEDVLLGEGAEQFLADDTWLEERFSDSAFIYA 120
| | | | |
DB 61 VERACFMSHAIVLEQALDEGLPIYIAVEEDVLLGEGAEQFLADDTWLEERFSDSAFIYA 120

QY 121 LETMFAKVIYRPDKVYNTENRSPFLLESEHCAGTIIISREARFPLDRFVAVLPERRIKA 180
| | | | |
DB 121 LETMFAKVIYRPDKVYNTENRSPFLLESEHCAGTIIISREARFPLDRFVAVLPERRIKA 180

QY 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSJK 240
| | | | |
DB 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSJK 240

QY 241 VMDLKRALGKFGREKKRMRORQAELEKVGRRVILFK 280
|||||
Db 241 VMDLKRALGKFGREKKRMRORQAELEKVGRRVILFK 280

RESULT 6
US-09-333-412-6
; Sequence 6, Application US/09333412
; Patent No. 6342382
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/333,412
; FILING DATE: 15-Jun-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: July 7, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-333-412-6

Query Match 100.0%; Score 1440; DB 4; Length 280;
Best local Similarity 100.0%; Pred. No. 7e-155;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MONHVISLASAERRAHIADTFGSRGIPQFDALMPSERLQAMAEVGLSAHPYLSG 60
|||||
Db 1 MONHVISLASAERRAHIADTFGSRGIPQFDALMPSERLQAMAEVGLSAHPYLSG 60
QY 61 VERACFMSHAVLWEQALDEGLPYIAVFEDVLLGEGAEQFLADDTWLEEFDDSAFIVR 120
|||||
Db 61 VERACFMSHAVLWEQALDEGLPYIAVFEDVLLGEGAEQFLADDTWLEEFDDSAFIVR 120
QY 121 LETMFAKVIYRPKVLNTEKRSFPLLESEHCGTAGYIISREARFFLDRAVLPPERIK 180
|||||
Db 121 LETMFAKVIYRPKVLNTEKRSFPLLESEHCGTAGYIISREARFFLDRAVLPPERIK 180
QY 181 VDLMEFTYFDKSGMPYQVSPALCTQELHYAKFLSONSMVGSDELEKDRQGRHRSLSK 240
|||||
Db 181 VDLMEFTYFDKSGMPYQVSPALCTQELHYAKFLSONSMVGSDELEKDRQGRHRSLSK 240
QY 241 VMDLKRALGKFGREKKRMRORQAELEKVGRRVILFK 280
|||||

Db 241 VMDLKRALGKFGREKKRMRORQAELEKVGRRVILFK 280

RESULT 7
US-09-338-943-6
; Sequence 6, Application US/09338943
; Patent No. 6379933
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: ROTH, STEPHEN
; APPLICANT: BUCZALA, STEPHANIE L.
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/338,943
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/478,140
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Cortuzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-338-943-6

Query Match 100.0%; Score 1440; DB 4; Length 280;
Best local Similarity 100.0%; Pred. No. 7e-155;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MONHVISLASAERRAHIADTFGSRGIPQFDALMPSERLQAMAEVGLSAHPYLSG 60
|||||
Db 1 MONHVISLASAERRAHIADTFGSRGIPQFDALMPSERLQAMAEVGLSAHPYLSG 60
QY 61 VERACFMSHAVLWEQALDEGLPYIAVFEDVLLGEGAEQFLADDTWLEEFDDSAFIVR 120
|||||
Db 61 VERACFMSHAVLWEQALDEGLPYIAVFEDVLLGEGAEQFLADDTWLEEFDDSAFIVR 120
QY 121 LETMFAKVIYRPKVLNTEKRSFPLLESEHCGTAGYIISREARFFLDRAVLPPERIK 180
|||||
Db 121 LETMFAKVIYRPKVLNTEKRSFPLLESEHCGTAGYIISREARFFLDRAVLPPERIK 180
QY 181 VDLMEFTYFDKSGMPYQVSPALCTQELHYAKFLSONSMVGSDELEKDRQGRHRSLSK 240
|||||
Db 181 VDLMEFTYFDKSGMPYQVSPALCTQELHYAKFLSONSMVGSDELEKDRQGRHRSLSK 240
QY 241 VMDLKRALGKFGREKKRMRORQAELEKVGRRVILFK 280
|||||

Db 241 VMEIDLKALGKFGREKKRMEORQAELEKYGRVILFK 280

RESULT 8

US-08-312-387B-8

; Sequence 8, Application US/08312387B

; Patent No. 5545553

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/312.387B

; FILING DATE: July 7, 1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-095

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 279 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-312-387B-8

Query Match 72.0%; Score 1036.5; DB 1; Length 279;

Best Local Similarity 71.6%; Pred. No. 3.5e-109;

Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;

QY 1 MONHVISLASAERAHAHTADTFGSRGIPFOFPDALMPSERLQAMAEIVPGLSAHPYLSG 60

DB 1 MONHVISLASAERAHAHTADTFGSRGIPFOFPDALMPSERLQAMAEIVPGLSAHPYLSG 60

QY 61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAEQFLAEDTWLQERFPDPSAFYVR 120

DB 61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAEQFLAEDTWLQERFPDPSAFYVR 120

QY 121 LETMFAYIVRPDKYLVNTENRSFPLESEHCSTAGYIISREARFELDFAVLPPERIKA 180

DB 121 LETMFAYIVRPDKYLVNTENRSFPLESEHCSTAGYIISREARFELDFAVLPPERIKA 180

QY 181 VDLAMFTYFEDEKGMPIYOVSPALCTQELHYAKFLSÖNSMLGSDLEKDEQGR----- 233

DB 181 VDLAMFTYFEDEKGMPIYOVSPALCTQELHYAKFLSÖNSMLGSDLEKDEQGR----- 233

QY 234 -----RHRSRLKVMFMDLKRALKFGREKKRMEORQAELEKYGRVILFK 280

DB 241 PANTFKHR-----LIRALTKIGRERKKRÖRR-----EOLIGKIIVPFO 279

RESULT 9

US-08-683-426-8

; Sequence 8, Application US/08683426

; Patent No. 5705367

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/683.426

; FILING DATE:

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/312.387

; FILING DATE: September 26, 1994

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-095B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 279 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-683-426-8

Query Match 72.0%; Score 1036.5; DB 1; Length 279;

Best Local Similarity 71.6%; Pred. No. 3.5e-109;

Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;

QY 1 MONHVISLASAERAHAHTADTFGSRGIPFOFPDALMPSERLQAMAEIVPGLSAHPYLSG 60

DB 1 MONHVISLASAERAHAHTADTFGSRGIPFOFPDALMPSERLQAMAEIVPGLSAHPYLSG 60

QY 61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAEQFLAEDTWLQERFPDPSAFYVR 120

DB 61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAEQFLAEDTWLQERFPDPSAFYVR 120

QY 121 LETMFAYIVRPDKYLVNTENRSFPLESEHCSTAGYIISREARFELDFAVLPPERIKA 180

DB 121 LETMFAYIVRPDKYLVNTENRSFPLESEHCSTAGYIISREARFELDFAVLPPERIKA 180

QY 181 VDLAMFTYFEDEKGMPIYOVSPALCTQELHYAKFLSÖNSMLGSDLEKDEQGR----- 233

DB 181 VDLAMFTYFEDEKGMPIYOVSPALCTQELHYAKFLSÖNSMLGSDLEKDEQGR----- 233

QY 234 -----RHRSRLKVMFMDLKRALKFGREKKRMEORQAELEKYGRVILFK 280

DB 241 PANTFKHR-----LIRALTKIGRERKKRÖRR-----EOLIGKIIVPFO 279

RESULT 10

US-08-683-458-8

; Sequence 8, Application US/08683458

; Patent No. 5798233

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-333-412-8

Query Match 72.0%; Score 1036.5; DB 4; Length 279;
Best Local Similarity 71.6%; Pred. No. 3,3e-109;
Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;

1 MONHYISLASAERRAHADTFGSGRIPPOFPDAMPSERLBOAMAEVPGLSAHPYLSG 60
1 MONHYISLASAERRAHADTFGSGRIPPOFPDAMPSERLBOAMAEVPGLSAHPYLSG 60

61 VERACPMASHVIMEQALDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
61 VERACPMASHVIMEQALDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

61 VERACPMASHVIMEQALDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
61 VERACPMASHVIMEQALDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

121 LETMAKVIYVRADKLVNENRSPFLSEHCHCTAGYIISREMRFPFLDFAVLPPERIKR 180
121 LETMAKVIYVRADKLVNENRSPFLSEHCHCTAGYIISREMRFPFLDFAVLPPERIKR 180

121 LETMAKVIYVRADKLVNENRSPFLSEHCHCTAGYIISREMRFPFLDFAVLPPERIKR 180
121 LETMAKVIYVRADKLVNENRSPFLSEHCHCTAGYIISREMRFPFLDFAVLPPERIKR 180

181 VDLMMETFEDEGMPYVSPALCTOELHYAKPLSONSMGLSGDLEKDEKDEGR----- 233
181 VDLMMETFEDEGMPYVSPALCTOELHYAKPLSONSMGLSGDLEKDEKDEGR----- 233

181 VDLMMETFEDEGMPYVSPALCTOELHYAKPLSONSMGLSGDLEKDEKDEGR----- 233
181 VDLMMETFEDEGMPYVSPALCTOELHYAKPLSONSMGLSGDLEKDEKDEGR----- 233

234 -----RHRSLSKVMFDLKRALGKFGKKRMRORQAELEKVGRRVILFK 280
234 -----RHRSLSKVMFDLKRALGKFGKKRMRORQAELEKVGRRVILFK 280

241 PANTFKR-----LIRALFKIGREKREKRORR-----EQLGKIIIVFQ 279
241 PANTFKR-----LIRALFKIGREKREKRORR-----EQLGKIIIVFQ 279

RESULT 13
US-08-646-590B-40
Sequence 40, Application US/08646590B
Patent No. 5962283
GENERAL INFORMATION:
APPLICANT: Warren, Patrick V.
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA

COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,590B
FILING DATE: 08-May-1996
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/599,171
FILING DATE: 09-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01094
FILING DATE: 21-January-1997

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/017001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal

US-08-646-590B-40

Query Match 6.9%; Score 99; DB 2; Length 394;
Best Local Similarity 22.8%; Pred. No. 0.0084;
Matches 75; Conservative 38; Mismatches 126; Indels 90; Gaps 17;

9 ASAAERRAHIAATFG-----SRGIP---PQFPA 34
9 ASAAERRAHIAATFG-----SRGIP---PQFPA 34

22 AKAKELRAKGVYVIFGAGEPDPDPDIFKACITALEGKTKYAPSACIPRLRAIAEK 81
22 AKAKELRAKGVYVIFGAGEPDPDPDIFKACITALEGKTKYAPSACIPRLRAIAEK 81

35 LMPSERLBOAMAEVPGLSAHPYLSGVEKACPMASHVIMEQALDEG-----LPYIAVEE 88
35 LMPSERLBOAMAEVPGLSAHPYLSGVEKACPMASHVIMEQALDEG-----LPYIAVEE 88

82 LKKEKVEKPESEIYVS-----AGAKMVLPL-----IFMALDGDDEVLLSPYVYTP 130
82 LKKEKVEKPESEIYVS-----AGAKMVLPL-----IFMALDGDDEVLLSPYVYTP 130

89 DVLGEGAEQFLAEDTWLEERFDKDSAFIVRLETFMAKVIYVRPK-VIN-----Y 138
89 DVLGEGAEQFLAEDTWLEERFDKDSAFIVRLETFMAKVIYVRPK-VIN-----Y 138

131 EQIRFEGV-----VEVPLKKEKGFOLSELVKEKVTERTKAIYVINSNNPTGAVY 182
131 EQIRFEGV-----VEVPLKKEKGFOLSELVKEKVTERTKAIYVINSNNPTGAVY 182

139 ENRSPFLSEHCHCTAG-YIISREMRFPFLDFAVLPPERIKAVDLMFTY--FPDKE 193
139 ENRSPFLSEHCHCTAG-YIISREMRFPFLDFAVLPPERIKAVDLMFTY--FPDKE 193

183 EEEELKKT-AEPCVERGIFITISDECYEYFVGDAKFVSPASDSDEKNTTFVNAFSSSY 241
183 EEEELKKT-AEPCVERGIFITISDECYEYFVGDAKFVSPASDSDEKNTTFVNAFSSSY 241

194 GMPYVQVSPALCTOELHYAKPL-SONSMGLSGDLEKDEKDEGRHRSLSKVMFDLKRALGK- 251
194 GMPYVQVSPALCTOELHYAKPL-SONSMGLSGDLEKDEKDEGRHRSLSKVMFDLKRALGK- 251

242 SMTGRRIGYVACPEE--YAKVIASLSNSQSVNVTTFQAQY-----ALEALKNPKSD 291
242 SMTGRRIGYVACPEE--YAKVIASLSNSQSVNVTTFQAQY-----ALEALKNPKSD 291

252 FGREKKRMRORQAELEKVGRRVILFK 277
252 FGREKKRMRORQAELEKVGRRVILFK 277

292 FVNEKRNMFERRRDYAVELSKIPGMDVY 320
292 FVNEKRNMFERRRDYAVELSKIPGMDVY 320

RESULT 14
US-09-412-184-40
Sequence 40, Application US/09412184
Patent No. 6268188
GENERAL INFORMATION:
APPLICANT: Warren, Patrick V.
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:

```

ADDRESS: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/412,184
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/646,590
FILING DATE: 08-May-1996
APPLICATION NUMBER: 08/599,171
FILING DATE: 09-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01094
FILING DATE: 21-January-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/017001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-412-184-40

```

```

Query Match          6.9%; Score 99; DB 4; Length 394;
Best Local Similarity 22.8%; Pred. No. 0.0084;
Matches 75; Conservative 38; Mismatches 126; Indels 90; Gaps 17;

QY 9 ASAAERRAHADTFG-----SRGIP---FQGFDA 34
DB 22 AKKKEIRAKGVYDIGAGCPDPTDFIKKACIRALRECKTYKASAGIPELRKIAEK 81
QY 35 LMESERLEQAMAEVPGLSAHAPYLSGVERKACFMSHAVALMEQALDEG-----LPYIAVEE 88
DB 82 LKKNKVEKPSSEIVS-----AGAKMYLFL---IFMILDEGDEVLLPSPYVWTYP 130
QY 89 DDLVLGEAGAEQFLAEDTWLEERDKSAFIVLETFPAKYIYAPDK-VLN-----Y 138
DB 131 EQIRFGGVP-----EVLPLKKEKGFQLSLQEDVKEVTEKTAIVNSPNNPTGAVY 182
QY 139 ENSFPLESEHCSTAG-YIISREARFPL--DRFAVLPPERIKAVDLMWTFV--FPDKE 193
DB 183 EEEELKRTI-AEFCVERGITIISDECEYTFYGDAKFVSPASSDEKYNITFTYNAFSKY 241
QY 194 GMEYVSPALCTQELHYAKFL-SQNSMLGSLDEKDEGRRRHRSLSKWFDLKRALGK- 251
DB 242 SMWGIRIGVACPEE--YAKVIASLNSQSVSNVTTFAYQG-----ALEALKNKNSMD 291
QY 252 FGREKKRMRORQA--ELEKYYGRRVI 277
DB 292 FVEMENAEFRDRDTAVEELSKIPGMDVY 320

```

```

RESULT 15
US-08-468-576B-17
; Sequence 17, Application US/08468576B
; Patent No. 5955345

```

```

GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,576B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kuit G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-468-576B-17

```

```

Query Match          5.8%; Score 83; DB 2; Length 712;
Best Local Similarity 20.2%; Pred. No. 1.3;
Matches 66; Conservative 59; Mismatches 106; Indels 96; Gaps 17;

QY 5 VISLSAERRAHADTFGSRGIPFQGFDAIMPSERLQAMAEVPGLSAHAPYLSG----- 60
DB 10 VADVLSQDLRKIKVOTFMDRCI-----LDLTKTQLR-NQLIHLELM-----HVLGSELQ 59
QY 61 ----VEKACFM--SHAVIMEQALDEGLPY-IAVEEDVVLGEAGAEQFLAEDTWLEERPD 112
DB 60 RSIISVEGSSLLIGASNSLVAHDLRGCGEYSLSVF-----FPESGLA 101
QY 113 KDSAFIVR-----LETMPAKYIYAPDKVLYNENRSPFL----- 145
DB 102 KERVFTMODLLQIKINPTSSLYKSLVSGSDK-----ENOKGFLMHFLKELAEYHOAKESC 157
QY 146 -LESEHCSTAGYIISREARFPLDRFAVLPPERIK--AVDLMWTFYFPDKEGMPYOVSP 202
DB 158 NMEIOTSSFTFNDSLAERKQLDDQFADAYPQRIKFESELEIKLYNKKRIE-----EQLR 213
QY 203 ALCTQELHYAKFLSONSMLGSLDEKDEGRRRHRSLSKWF--DLKRALGKGF-----RE 255
DB 214 EMC-QIKKPFK-----DTEIAKIKMEAKKKYKEKELTMFQNDPEKACQAKSEALVIRE 264

```

OY 256 KK-KRMEQQAEELEKVGRRVILEK 280
| : : : | : : | : |
DB 265 KSTLERIHKHQEIETKTEIYAQRQLLK 291

Search completed: December 2, 2002, 12:03:18
Job time : 13.4223 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 12:00:15 ; Search time 5.61203 Seconds
(without alignments)
794.504 Million cell updates/sec

Title: US-10-007-267-6

Perfect score: 1440
Sequence: 1 MGNHVISLASAERRAHAD.....ERQRAELEKVGRRVILFK 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/PC1_NEW_PUB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubpaa/PC1US_PUBCOMB.pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*

12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1440	100.0	280	US-10-007-267-6	Sequence 6, Appl1
2	1036.5	72.0	279	US-10-007-267-8	Sequence 8, Appl1
3	1007	69.9	275	US-09-211-691-2	Sequence 2, Appl1
4	99	6.9	394	US-09-905-173-40	Sequence 40, Appl1
5	87.5	6.1	657	US-09-815-242-5280	Sequence 5280, Ap
6	87.5	6.1	662	US-09-815-242-12140	Sequence 12140, A
7	85	5.9	762	US-09-815-242-11519	Sequence 11519, A
8	83.5	5.8	856	US-09-815-242-11310	Sequence 11310, A
9	79	5.5	2055	US-10-017-216-4	Sequence 4, Appl1
10	77.5	5.4	222	US-09-864-761-48086	Sequence 48086, A
11	77.5	5.4	363	US-09-849-031A-1	Sequence 1, Appl1
12	77.5	5.4	363	US-09-849-562A-1	Sequence 1, Appl1
13	77.5	5.4	623	US-09-815-242-13499	Sequence 13499, A
14	77.5	5.4	1224	US-09-801-368-222	Sequence 222, App
15	77.5	5.4	4999	US-09-861-059-14	Sequence 14, Appl1
16	77	5.3	592	US-09-861-451A-72	Sequence 72, Appl1
17	76.5	5.3	856	US-09-815-242-11489	Sequence 11489, A
18	76.5	5.3	916	US-09-815-242-1174	Sequence 1174, App
19	76.5	5.3	1005	US-09-925-301-1335	Sequence 1335, Ap

20	75.5	5.2	989	US-09-815-242-4897	Sequence 4897, Ap
21	74	5.1	302	US-09-815-242-5371	Sequence 5371, Ap
22	74	5.1	325	US-09-815-242-12608	Sequence 12608, A
23	73.5	5.1	883	US-09-815-242-13382	Sequence 13382, A
24	73.5	5.1	1427	US-09-991-496-97	Sequence 97, Appl1
25	73.5	5.1	1427	US-09-874-923-97	Sequence 97, Appl1
26	73.5	5.1	1641	US-09-991-496-96	Sequence 96, Appl1
27	73.5	5.1	1641	US-09-874-923-96	Sequence 96, Appl1
28	73	5.1	474	US-09-815-242-5389	Sequence 5389, A
29	73	5.1	474	US-09-815-242-12656	Sequence 12656, A
30	73	5.1	3014	US-09-737-149-2	Sequence 2, Appl1
31	72.5	5.0	556	US-09-887-586A-32	Sequence 32, Appl1
32	72.5	5.0	556	US-09-903-012-32	Sequence 32, Appl1
33	72.5	5.0	1237	US-09-841-132-592	Sequence 592, App
34	72.5	5.0	1579	US-09-801-368-368	Sequence 368, App
35	72	5.0	464	US-09-902-941-1906	Sequence 1906, Ap
36	72	5.0	468	US-09-925-300-1620	Sequence 1620, Ap
37	72	5.0	793	US-09-815-242-13689	Sequence 13689, A
38	72	5.0	1356	US-09-801-368-306	Sequence 306, App
39	71.5	5.0	99	US-09-864-761-33635	Sequence 33635, A
40	71.5	5.0	552	US-09-817-764-4	Sequence 4, Appl1
41	71.5	5.0	883	US-09-815-242-13684	Sequence 13684, A
42	71	4.9	420	US-09-844-006A-2	Sequence 2, Appl1
43	71	4.9	710	US-09-815-242-10895	Sequence 10895, A
44	71	4.9	734	US-09-764-367A-4	Sequence 4, Appl1
45	71	4.9	893	US-09-916-790-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-10-007-267-6
Sequence 6, Application US/10007267
Patent No. US20020127682A1

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF OLIGOSACCHARIDES, AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 280 amino acids
TYPE: amino acid

```
;
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-007-267-6

Query Match          100.0%; Score 1440; DB 12; Length 280;
Best Local Similarity 100.0%; Pred. No. 2.3e-138;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGNHVISLASAARRAHIAIDTFGSRGIPFOFPALMPSERLBOAMALVGLSAHPYLSG 60
DB 1 MGNHVISLASAARRAHIAIDTFGSRGIPFOFPALMPSERLBOAMALVGLSAHPYLSG 60
QY 61 VERACFMSHAIVLWEQALDEGLPIYIAVEDDVLLEGAEQFLADDTMLEERFDKDSAFYR 120
DB 61 VERACFMSHAIVLWEQALDEGLPIYIAVEDDVLLEGAEQFLADDTMLEERFDKDSAFYR 120
QY 121 LETMFAKVIYRPDKVLNENRSPFLLESEHCAGTGYIISREAMRFLDRPAVLPPERIKA 180
DB 121 LETMFAKVIYRPDKVLNENRSPFLLESEHCAGTGYIISREAMRFLDRPAVLPPERIKA 180
QY 181 VDLMMFTYFDKGMPIYQVSPALCTQELHYAKFLSONSMGLSDLEKDRQGRHRRLSK 240
DB 181 VDLMMFTYFDKGMPIYQVSPALCTQELHYAKFLSONSMGLSDLEKDRQGRHRRLSK 240
QY 241 VMFDLKRALGKFGREKKRMRERQROAELEKYGRVILFK 280
DB 241 VMFDLKRALGKFGREKKRMRERQROAELEKYGRVILFK 280

RESULT 2
US-10-007-267-8
; Sequence 8, Application US/10007267
; Patent No. US20020127682A1
; GENERAL INFORMATION:
;   APPLICANT: Gotschlich, Emil C.
;   TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;                     OLIGOSACCHARIDES, AND GENES ENCODING THEM
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Klauber & Jackson
;     STREET: 411 Hackensack Avenue
;     CITY: Hackensack
;     STATE: New Jersey
;     COUNTRY: USA
;     ZIP: 07601
;   COMPUTER READABLE FORM:
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/10/007,267
;     FILING DATE: 03-Dec-2001
;     CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/09/333,412
;     FILING DATE: 15-Jun-1999
;     APPLICATION NUMBER: 08/312,387
;     FILING DATE: July 7, 1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Jackson Esq., David A.
;     REGISTRATION NUMBER: 26,742
;     REFERENCE/DOCKET NUMBER: 600-1-095
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 201 487-5800
;     TELEFAX: 201 343-1684
;     TELEX: 133521
;   INFORMATION FOR SEQ ID NO: 8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 279 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
```

```
;
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-007-267-8

Query Match          72.0%; Score 1036.5; DB 12; Length 279;
Best Local Similarity 71.6%; Pred. No. 1.6e-97;
Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;

QY 1 MGNHVISLASAARRAHIAIDTFGSRGIPFOFPALMPSERLBOAMALVGLSAHPYLSG 60
DB 1 MGNHVISLASAARRAHIAIDTFGSRGIPFOFPALMPSERLBOAMALVGLSAHPYLSG 60
QY 61 VERACFMSHAIVLWEQALDEGLPIYIAVEDDVLLEGAEQFLADDTMLEERFDKDSAFYR 120
DB 61 VERACFMSHAIVLWEQALDEGLPIYIAVEDDVLLEGAEQFLADDTMLEERFDKDSAFYR 120
QY 121 LETMFAKVIYRPDKVLNENRSPFLLESEHCAGTGYIISREAMRFLDRPAVLPPERIKA 180
DB 121 LETMFAKVIYRPDKVLNENRSPFLLESEHCAGTGYIISREAMRFLDRPAVLPPERIKA 180
QY 181 VDLMMFTYFDKGMPIYQVSPALCTQELHYAKFLSONSMGLSDLEKDRQGR----- 233
DB 181 VDLMMFTYFDKGMPIYQVSPALCTQELHYAKFLSONSMGLSDLEKDRQGR----- 233
QY 234 -----RHRRSLKVMFDLKRALGKFGREKKRMRERQROAELEKYGRVILFK 280
DB 241 PAMTFKHR-----LIRALTKIGREKRKROR-----EQLIGKIYVEFQ 279

RESULT 3
US-09-211-691-2
; Sequence 2, Application US/09211691
; Patent No. US20020034805A1
; GENERAL INFORMATION:
;   APPLICANT: Gilbert, Michel
;   APPLICANT: Young, N. Martin
;   APPLICANT: Wakarchuk, Warren W.
;   TITLE OF INVENTION: National Research Council of Canada
;   TITLE OF INVENTION: Fusion Proteins for Use in Enzymatic Synthesis of
;   FILE REFERENCE: 019957-012910US
;   CURRENT APPLICATION NUMBER: US/09/211,691
;   CURRENT FILING DATE: 1998-12-14
;   PRIOR APPLICATION NUMBER: US 60/069,443
;   PRIOR FILING DATE: 1997-12-15
;   NUMBER OF SEQ ID NOS: 18
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 2
;   LENGTH: 275
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-211-691-2

Query Match          69.9%; Score 1007; DB 10; Length 275;
Best Local Similarity 74.7%; Pred. No. 1.5e-94;
Matches 201; Conservative 20; Mismatches 44; Indels 4; Gaps 1;

QY 1 MGNHVISLASAARRAHIAIDTFGSRGIPFOFPALMPSERLBOAMALVGLSAHPYLSG 60
DB 1 MGNHVISLASAARRAHIAIDTFGSRGIPFOFPALMPSERLBOAMALVGLSAHPYLSG 60
QY 61 VERACFMSHAIVLWEQALDEGLPIYIAVEDDVLLEGAEQFLADDTMLEERFDKDSAFYR 120
DB 61 VERACFMSHAIVLWEQALDEGLPIYIAVEDDVLLEGAEQFLADDTMLEERFDKDSAFYR 120
QY 121 LETMFAKVIYRPDKVLNENRSPFLLESEHCAGTGYIISREAMRFLDRPAVLPPERIKA 180
DB 121 LETMFAKVIYRPDKVLNENRSPFLLESEHCAGTGYIISREAMRFLDRPAVLPPERIKA 180
QY 181 VDLMMFTYFDKGMPIYQVSPALCTQELHYAKFLSONSMGLSDLEKDRQGR----- 236
DB 181 VDLMMFTYFDKGMPIYQVSPALCTQELHYAKFLSONSMGLSDLEKDRQGR----- 240
```

OY 237 RSLKVFEDIKRALGFGREKKRMRERQ 265
 Db 241 PANTFKHLRALTKISRERKRORREQ 269

RESULT 4

US-09-905-173-40
 ; Sequence 40, Application US/09905173
 ; Patent No. US2002013295A1

GENERAL INFORMATION:

APPLICANT: DIVERSA CORPORATION
 APPLICANT: SHORT, Jay M.
 APPLICANT: WARREN, Patrick V.
 APPLICANT: SWANSON, Ronald V.
 APPLICANT: MATHUR, Eric J.
 TITLE OF INVENTION: ENZYMES HAVING TRANSAMINASE AND AMINOTRANSFERASE ACTIVITY AND MET
 FILE OF INVENTION: USE THEREOF

FILE REFERENCE: DIVER1240-7
 ; CURRENT APPLICATION NUMBER: US/09/905,173
 ; CURRENT FILING DATE: 2001-07-12
 ; PRIOR APPLICATION NUMBER: US 09/412,184
 ; PRIOR FILING DATE: 1999-10-04
 ; PRIOR APPLICATION NUMBER: US 09/389,537
 ; PRIOR FILING DATE: 1999-09-02
 ; PRIOR APPLICATION NUMBER: US 08/646,590
 ; PRIOR FILING DATE: 1996-05-08
 ; PRIOR APPLICATION NUMBER: US 08/599,171
 ; PRIOR FILING DATE: 1996-02-09
 ; PRIOR APPLICATION NUMBER: US 09/481,733
 ; PRIOR FILING DATE: 2000-01-11
 ; PRIOR APPLICATION NUMBER: US 09/069,226
 ; PRIOR FILING DATE: 1998-04-27
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 40
 ; LENGTH: 394
 ; TYPE: PRT
 ; ORGANISM: Aquifex
 ; US-09-905-173-40

Query Match 6.9%; Score 99; DB 10; Length 394;
 Best Local Similarity 22.8%; Pred. No. 0.021;
 Matches 75; Conservative 38; Mismatches 126; Indels 90; Gaps 17;

OY 9 ASAAERRAHIIADPFG-----SRGIP---FQFPDA 34
 Db 22 AAKAKELRAGGVYIGFGAGEPDPTDFPKKACIRALREGKTKYAPSAQIPELREAIK 81
 OY 35 LMPSERLQAMAEIVPGLSAHPYLSGVEKACPMASHAVLMEQALDEG-----LPYIAVEE 88
 Db 82 LTKENKVEYKPSSEIVS-----AGAKMVLFL---IFMILDEGDEVLLSPYWTYP 130
 OY 89 DQVLGEGAEQFLAEDTWMLEERFDKDSAFIVRLTETMAFVIYAPDK-VLN-----Y 138
 Db 131 EDIRFEGVP-----VEVPLKKEKGFQLSLEDVKEVTEKTAIVINSPNPTGAVY 182
 OY 139 ENRSFPLESEHOGTAG-YIISREARFPL--DRFAVLPPERIKAVDLAMFTY--FFDKE 193
 Db 183 EEBELKKT-AEFCVGERGITIISDECEYFVYGDAKFVSASDSDEKNTITFYNAFSKY 241
 OY 194 GMPYQVSPALCTQELHYAKFL-SQNSMLGSDLEKDEQGRHRRSLKWMFDIKRALGK- 251
 Db 242 SMTGMRIGVACPEE--YAKVIASLNSQSVNVTTPAOG-----ALEALKNPKSKD 291
 OY 252 FGREKKRMRERQRA---ELEKYYGRRVI 277
 Db 292 FVEMRNARERDRDTAVEELSKIPGMDVY 320

RESULT 5
 US-09-815-242-5280
 ; Sequence 5280, Application US/09815242
 ; Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes In
 TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5280
 ; LENGTH: 657
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 ; US-09-815-242-5280

Query Match 6.1%; Score 87.5; DB 10; Length 657;
 Best Local Similarity 23.1%; Pred. No. 0.61;
 Matches 54; Conservative 22; Mismatches 79; Indels 79; Gaps 7;

OY 29 FQFPDALMPSERLQAMA-----ELVGLSAHPYLSGVEA 64
 Db 429 FVFSDLKPKALRLSSIMGNATFTFTHDSIANGEDGPTHEPIQLAGLRAIPMNVIRPA 488
 OY 65 CFMSHAVLMEQALD-----EGLPYIAVEEDV----- 91
 Db 489 DGNETVYVMEVALLSESTPSTSLVLTNRNLPVLDPEDVYBEGVRKCAITYYGSSETPFL 548
 OY 92 LLEGAEQFLAEDTWMLEERFDKDSAFIVRLTETMA-----KVIYRPDKVLNVE--NR 141
 Db 549 LTAGSSEVSLAVEAKDLKQKSVYVSPMNNAPFEQSEKESVIPSSTKRVAIEM 608
 OY 142 SFPLESEHCAGTAGYIISREARFPLDRFAVLPPERIKAVDLAMFTYFPDKGEM 195
 Db 609 ASPLGMHKYVGTAGKVIA-----IDGFCASAPG-----DLAVEKYYGTTKEMI 650

RESULT 6
 US-09-815-242-12140
 ; Sequence 12140, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes In
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A

```

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 12140
; LENGTH: 662
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12140

Query Match          6.1%; Score 87.5; DB 10; Length 662;
Best Local Similarity 23.1%; Pred. No. 0.62;
Matches 54; Conservative 22; Mismatches 79; Indels 79; Gaps 7;

QY 29 FGFPLALMSESLQAMA-----ELVGLAHPLSLVEKA 64
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 434 FVFSDLKPLALRISSTIGLNATFTHTDSIAGEGDPTEHPLEQLALRAIPMNVIRBA 493
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 65 CFMSHAVLWEQALD-----EGLPYIAVEEDV----- 91
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 494 DGNETVNAEVALESSTPTSLVLTQNLPLVDVPEDEVVEGVRKATYVYSEETPEFL 553
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 92 LAGEAGQFLAEDPTLEERFDKDSAFIVRLTFMA-----KVIYRDKVLNVE--NR 141
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 554 LLASGEVSLAVEAAKDLCKGKSVRVGMPMNAFEQOSEEKESVYKRVAIM 613
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 142 SPPLESEHCAGAGYIISREARMFILDRFAVLPPERIKAVDLMFTYFDDKEM 195
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 614 ASPLGWHKTYGTAGKYIA-----IDFGASAPG-----DLVERKYGTPTKEMI 655
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 7
US-09-815-242-11519
; Sequence 11519, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11519
; LENGTH: 762
; TYPE: PRF
; ORGANISM: Helicobacter pylori
US-09-815-242-11519

Query Match          5.9%; Score 85; DB 10; Length 762;
Best Local Similarity 25.0%; Pred. No. 1.3;
Matches 53; Conservative 32; Mismatches 65; Indels 62; Gaps 12;

QY 89 DQVLAGEAGQFLA-EDTLEERFDKDSAFIV-----RLTFPAKYIVPDVAVNTENS 142
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 412 DETELGTDADEASSLYKTLLEKLKONQIVITTHKRSLVMA-----ENKE 459
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 143 FPLL-----ESBH-----CGTAGYIISRE-AMREFLDRAVLPPERIKAVDLMFTYF 189
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 460 VELLAALYDEEKRPYTYTLKGYIGKSYAFETALRGVFPFLI--EKAKAF-----YG 510
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 190 FKEGMPYVQVSPALCTQELHYAKFLSQNSMLGSDLEKDRQGRH-RSLKVMEDLKRA 248
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 511 EDKEKLVN-----LIENS-----STLERELKQNEHLNALKEQEDLKNA 550
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 249 LKFGREKKRMRQQALEKVGRRVILFK 280
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 551 -WLEMEKQKEIFHHRKLEKRSYQOALNLIK 581
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 8
US-09-815-242-11310
; Sequence 11310, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11310
; LENGTH: 856
; TYPE: PRF
; ORGANISM: Helicobacter pylori
US-09-815-242-11310
```

Query Match	5.8%	Score 83.5	DB 10	Length 856
Best Local Similarity	21.7%	Pred. No. 2.2		
Matches 54	Conservative 40	Mismatches 98	Indels 57	Gaps 12
QY	39	ERLQAMAEELVPGSAHFYLSGVKACFMSHAVLMEQALDEGLPIYAFVEDVLGE---	95	
Db	255	ERLKVIVIEVKKSNVILFIDEI-----HTIYGAGSEGGMDAANLTKALARGELHT	307	
QY	96	GABQFLAEDTWLEE---REDKDSAFIVRETFMFAKVIYPPDKVLNENRSPFL-----	145	
Db	308	IGA-----TTLKEKRYKFEKDMA---LORRQPIILNPSI---NEALQILRLGIKE	352	
QY	146	-LESEH---CCTAGYIISRE-AMRFPLDRFAVLPERIKADLMMFTYFPKDEGPPYQV	200	
Db	353	TELEHNTTINDSALLISAKLSRITYDR--LPKALIDLDEGAQLKQWSESPAKLS	410	
QY	201	SPALCTOELHAKFLSONSMIGSDLEKDEQRRHRSILKYMFLDKRALGKFGREKKRM	260	
Db	411	SVKSTIGRLMEK-----QALEMEKESNAKRNQIFLTKELSDLK-----EKKIQL	455	
QY	261	ERQROAELE 269		
Db	456	EAQFENEKE 464		
RESULT 9				
US-10-017-216-4				
Sequence 4	Application US/10017216			
Patent No.	US20020160483A1			
GENERAL INFORMATION:				
APPLICANT:	KAPELLER-LIBERMANN, Rosana			
TITLE OF INVENTION:	13245, A No. US20020160483A1el Human Myotonic Dystrophy Type Prod			
TITLE OF INVENTION:	Kinase and Uses Therefor			
FILE REFERENCE:	10147-5701			
CURRENT APPLICATION NUMBER:	US/10/017, 216			
PRIOR APPLICATION NUMBER:	US/2001-10-23			
PRIOR FILING DATE:	2000-10-23			
NUMBER OF SEQ ID NOS:	7			
SOFTWARE:	Patentin Ver. 2.1			
SEQ ID NO 4				
LENGTH:	2055			
TYPE:	PRT			
ORGANISM:	Mus musculus			
US-10-017-216-4				
Query Match	5.5%	Score 79	DB 9	Length 2055
Best Local Similarity	17.1%	Pred. No. 21		
Matches 49	Conservative 55	Mismatches 99	Indels 84	Gaps 12
QY	29	FOFF-----DALMPSERLQOAMAEI-----VGLSAHFYLSGVKACFMSHAVLMEQ	75	
Db	317	FQRLKFPDPDKVSELLDILQSLCVCQKERLKEGLOCHPEFARTD-----WNN	366	
QY	76	ALDEGLPIYAFVEDVLGEBAQDFLAEDTWLEERFDKDSAFIVRETFMFAKVIYPPDKV	135	
Db	367	IRNSPPPVPTLKS-----DDT---SNFDEPE-----KSMAFILCVPAEP	405	
QY	136	LNENRSPFLLESEHCTAGYIISREAMRFPLDRRAVLPERIKAVDLMF-----	186	
Db	406	LAFSEELLPVGFSTKALGYIGRSESVSSLD-----SPAKVSSMEKLLIKSELODS	460	
QY	187	---TYFPFKE-----GMPYQVSPALCTOELHAKFLSONSMIGSD-----LEKDR	229	
Db	461	ODKCKMKQDMTRLHRRRSEVEAVLSQKEVLKASERTQSLLEDQLATYITECSLKSL	520	
QY	230	EOGR-----RHRSLKVMFDKRALGKFGREKKRMERQROAELEKV 271		
Db	521	EQARMEVSEQDDKALQLLHDIREQ-----SRKLQETKEQEYQAVQVEEM 563		
RESULT 10				
US-09-864-761-48086				

[illegible]

APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 222
LENGTH: 1224
TYPE: PRF
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-222

Query Match 5.4%; Score 77.5; DB 10; Length 1224;
Best Local Similarity 18.6%; Pred. No. 15;
Matches 43; Conservative 47; Mismatches 90; Indels 51; Gaps 11;

QY 17 HIAIDFGSGRIGPFQFDALMPSERLQAMAEVPLSAPHYLSGVER---ACFMSHAVL 72
DB 479 YFSSPFGQVLSSTFLD-----HKLE-----PYLGALSQYMIVECFINGCIR 520
QY 73 W-----EGALDEGLPIYAFEDVDVLGSGAQFLAEDTWLEERFDKDSAFIVRL 122
DB 521 WKIWTGDDYDEKIDSLQKLEILSNOLIALN---LREPLLRKQIQONFALFTMLK 575
QY 123 TMEAKYIVPRDKVLYENRSPFLSEHCGTAGYIIS--REARFPLDRFAVLPPEKKA 180
DB 576 DNVLFLL--EKITSATMDYPRINEEBRGAESDAVRLRACGTELRNALIMPESLKK 633
QY 181 VDIAMTFYFDKGMFVYOVSPALCTQF--LHAKELSONSMIGS--DLEKDR 229
DB 634 I-----YPDLESV--IARIMPLSYHEKISFKSFLIITVLKSLDMKEER 676

RESULT 15
US-09-976-059-14
Sequence 14, Application US/09976059
Patent No. US20020164747A1
GENERAL INFORMATION:
APPLICANT: Farnet, Chris
APPLICANT: Zazopoulos, Emmanuel
APPLICANT: Stafia, Alfredo
TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
FILE REFERENCE: 3019-PCT
CURRENT APPLICATION NUMBER: US/09/976,059
CURRENT FILING DATE: 2001-10-15
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.0
SEQ ID NO 14
LENGTH: 4999
TYPE: PRF
ORGANISM: Actinoplanes sp.
US-09-976-059-14

Query Match 5.4%; Score 77.5; DB 9; Length 4999;
Best Local Similarity 21.8%; Pred. No. 96;
Matches 55; Conservative 38; Mismatches 100; Indels 59; Gaps 13;

QY 5 VISLASAARRRAHI-----ADTFGS---RGIPROFFDALMPSERLQAMAEVPLGISA 54
DB 2896 LVNMGITETTVYHQDLAPADTSSPIGRIGP-----LSVYVLDLALRPVPGVAG 2949
QY 55 HPIYLSGVERKACFMSHAVLMEQALD---EGLPIYANFE-----DDVLLGEGAEQFLAF- 103
DB 2950 EYVYVAGRQ---LARYLYLGRALITGRFVACPFPLPAGERMRYTGDARAKWSRGRLOPAGRT 3005
QY 104 DFWLE-----ERFDKDSAFIVRLTMAFAKIVIRPD-----KVLNENRSPFLSEHCG 152

DB 3006 DDQVQIRGFRIEBPGEVQAVVAHAHPEIAAAAVVREDVPDGPRLTAYVVPAGP-----R 3058
QY 153 TACYIISREARREFLDRF--AVLPPEKIAVDLMFTYF--FDKEGMPYIQ-----VSP 202
DB 3059 TAPAAVAETVRRFPADRLPAYMLPSAVVLDALPLTDHGKIDRRALPAPQHTGASGRAP 3118
QY 203 ALCTOELHYAKF 214
DB 3119 ATVAEEVLCNAF 3130

Search completed: December 2, 2002, 12:16:05
Job time : 8.61203 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 12.4266 Seconds
(without alignments)
2166.126 Million cell updates/sec

Title: US-10-007-267-6
Perfect score: 1440
Sequence: 1 MGNHVISLASAERRAHAD.....ERQQAELKVGRRVILFK 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1284	89.2	280	A81027	lacto-N-neotetraos
2	1236	85.8	276	S70815	glycosyl transferase
3	1012	70.3	275	C81027	lacto-N-neotetraos
4	1007	69.9	268	B81971	lacto-N-neotetraos
5	1007	69.9	275	S70814	glycosyl transferase
6	971.5	67.5	279	A81971	lacto-N-neotetraos
7	500.5	34.8	266	S71025	lipopolysaccharide
8	438.5	30.5	302	A64077	lipopolysaccharide
9	315	21.9	282	F64091	lipopolysaccharide
10	233	16.2	268	A83304	lipopolysaccharide
11	121	8.4	738	T00343	lipooligosaccharid
12	120.5	8.4	332	E71916	hypothetical prote
13	108.5	7.5	267	E64175	probable lipopolys
14	104	7.2	273	B64623	hypothetical prote
15	99	6.9	273	E71890	lipopolysaccharide
16	99	6.9	324	A70469	probable lipopolys
17	98.5	6.8	721	AH3417	aspartate transam
18	97	6.7	444	F71916	lipsa protein [limp
19	97	6.7	839	F85334	probable lipopolys
20	97	6.7	1446	T04528	myosin heavy chain
21	95.5	6.6	575	B83959	myosin heavy chain
22	95.5	6.6	1242	T45976	hypothetical prote
23	95	6.6	23	T20368	myosin heavy chain
24	93.5	6.5	284	E64620	lipopolysaccharide
25	92.5	6.4	404	C64597	lipopolysaccharide
26	92.5	6.4	457	T40770	colled coil protei
27	91	6.3	419	T19837	hypothetical prote
28	90	6.2	1953	T40642	probable helicase
29	89.5	6.2	266	AG2797	conserved hypothet

30	89.5	6.2	266	2	H97576	hypothetical prote
31	89.5	6.2	594	2	I49127	intracellular prot
32	88.5	6.1	256	2	T00097	hypothetical prote
33	88.5	6.1	885	1	VCPWF2	structural protein
34	88.5	6.1	2415	1	A37333	spectrin alpha cha
35	88	6.1	292	2	D71894	probable lipopolys
36	88	6.1	998	2	S47105	myosin heavy chain
37	87.5	6.1	314	2	G71974	probable keto-acid
38	87.5	6.1	429	2	A36220	transforming prote
39	87.5	6.1	662	2	G89909	transketolase [imp
40	87	6.0	468	1	D26190	phosphoglucosate d
41	87	6.0	2970	2	T08839	polyprotein - marm
42	86	6.0	291	2	AB1384	conserved hypothet
43	85.5	5.9	653	2	AB1128	transcription anti
44	85.5	5.9	1502	2	T14278	myosin-like protei
45	85.5	5.9	1529	2	A59310	unconventional myo

ALIGNMENTS

```
RESULT 1
A81027
lacto-N-neotetraose biosynthesis glycosyl transferase Igte NMB1926 [imported] - Neiss
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: A81027
R:Retellein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
rl, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiarant, V.; Pizsa, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappunli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; M01D:20175755; PMID:10710307
A:Accession: A81027
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-280 <TEP>
A:Cross-references: GB:AE002541; GB:AE002098; NID:g7227175; PIDN:AA42255.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1926
C:Superfamily: lipopolysaccharide biosynthesis-associated protein

Query Match      89.2%   Score 1284;   DB 2;   Length 280;
Best Local Similarity 90.4%   Pred. No. 3.3e-104;
Matches 253; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 1 MGNHVISLASAERRAHADTFGRGIPQFPDAMPSERLQAMAEIVPGLSAHPYLSG 60
    |||||||
DB 1 MGNHVISLASAERRAHADTFGRGIPQFPDAMPSERLQAMAEIVPGLSAHPYLSG 60
    |||||||

QY 61 VERACFMSHAVIMEQALDEGLPIYIAVEEDVILGEGAEQFLAEDTWLEERFDKDSAFIVR 120
    |||||||
DB 61 VERACFMSHAVIMEQALDEGLPIYIAVEEDVILGEGAEQFLAEDTWLEERFDKDSAFIVR 120
    |||||||

QY 121 LETMPKXIVRPDKVINYNSRSPILSEHCGTAGTISRMRPFLDFAVLPERRITA 180
    |||||||
DB 121 LETMPKXIVRPDKVINYNSRSPILSEHCGTAGTISRMRPFLDFAVLPERRITA 180
    |||||||

QY 181 VDLMEFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMLSDLEKDEQGRHRRSLK 240
    |||||||
DB 181 VDLMEFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMLSDLEKDEQGRHRRSLK 240
    |||||||

QY 241 VMFDLKRALGKFGREKKRMRERQQAELKVGRRVILFK 280
    |||||||
DB 241 VMFDLKRALGKFGREKKRMRERQQAELKVGRRVILFK 280
    |||||||

RESULT 2
S70815
glycosyl transferase E (EC 2.4.-.-) - Neisseria meningitidis
C:Species: Neisseria meningitidis
```


RESULT 11
T00343
hypothetical protein KIAA0584 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00343
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T00343
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-738 <NAG>
A:Cross-references: EMBL:AB011156; NID:g3043691; PIDN:BA25510.1; PID:g3043692
A:Experimental source: brain
C:Genetics:
A>Note: KIAA0584

Query Match 8.4%; Score 121; DB 2; Length 738;
Best Local Similarity 23.0%; Pred. No. 0.019;
Matches 61; Conservative 51; Mismatches 107; Indels 46; Gaps 15;

QY 5 VISASAAERAAHIAADTFGSGIPPFDAI---MPSELEQAMAEIVPG---LSAHP 56
Db MINIKRRDRDRMLRTLYEIEYKIVEADVQKALNTSQLKALNIEMLPGYRDPYSSRP 517
QY 57 YLSGVEKCFMSHAVLMEQALDEGLPIYAVFEDVLLGEGAEQFLAE--DWLIERFQKD 114
Db LTRG-ETICGFLSHSVKMEVVDREKTLVIEDVDFEHOFRKKIMKMDIDQALDWE 576
QY 115 SAFVRLKETMAKVIYRPDK-VLVYENRSFPLLESEHC-GTAGYLIISREARFEL--DRF 170
Db LIYIGRRKMQ---VKPEKAVPNVAN---LVADYSYWLGLGYISLEGAKLVGAMPF 628
QY 171 A-VLPPEIKAV-----DLMAFTTFDEKGMPIYQVSPALCTOELHYV---KFLS----- 216
Db 629 GKMLPVDFELPMYMKHAPVARYKEYESRDIAKFASEP-LIYPRHYGQPGYLSDFETS 687
QY 217 ---ONSMGSDLE-----KDEKGR 233
Db 688 TIMDNEYATIMDRHAKSKQSR 712

RESULT 12
E71916
probable lipopolysaccharide biosynthesis protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: E71916
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ires, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: E71916
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <KAN>
A:Cross-references: GB:AE001489; GB:AE001439; NID:94155102; PIDN:AA06135.1; PID:9415510
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0562

Query Match 8.4%; Score 120.5; DB 2; Length 332;
Best Local Similarity 22.6%; Pred. No. 0.0078;
Matches 79; Conservative 47; Mismatches 98; Indels 125; Gaps 21;
QY 4 HVISASAAERAAHIAADTFGSGIPPFDAI---MPSELEQAMAEIVPG---LSAHP 56
Db 5 YIISLKESQRR---LDTEKVLSENEKFGRCV-FQIFDAISPKHQDFEKLQELYNAAQ 59

QY 53 SA-----HPY---LSGVEKACFMSSHAVLMEQALDEGLPIYAVFEDVLLGEGAEQFLA 102
Db SLQSDWHSYVAGLTLPELGCYLSHYLMECKVKLDQ--VLIEDDYLTLESHMQAL- 117
QY 103 EDTWLEERFDKDSAFIVL-----ETMF---AKVIYRPDKVLYENRSFPLES--- 148
Db 118 EDC-LKSPFD---FVRLGYCYWYHETKPHVLPEKEVFPFDFHSKNN--PILEKPK 169
QY 149 -----EH-----CGTAGYLIISREA 162
Db 170 FFDVSRLNLSTHKYIHYILKKICKSYATHEKAFLEHFEYLSVYASTAGYILTFKQ 229
QY 163 MRFFL---DREAVLPERIKAVDLMAFTYFPDKESGMPYQVSPALCTOELHYAKFLQNS 219
Db 220 AKTFLEATESKILPE-----VDMFMDSAHVDVANLTYVCPV-----SLSEHS 274
QY 220 MGSDELKEDREGRR-----HRSKLYMF---DLKRALGKFGREKK 258
Db 275 -LDSTIQPKOKSKLSKYPDPQSTFKMLFYSLNAKKRLNAFOQYSKQ 322

RESULT 13
E64175
hypothetical protein H11697 (1sg locus) - Haemophilus influenzae
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: E64175; S27580
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Goearne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:99350630; PMID:7542800
A:Accession: E64175
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-267 <TIGR>
A:Cross-references: GB:I42023; TIGR:H11697
A:Experimental source: strain Rd KW20
R:McLaughlin, R.; Abu Kwaik, Y.; Young, R.; Spinoia, S.; Apicella, M.
submitted to the EMBL Data Library, June 1992
A:Description: Characterization and sequence of the 1sg locus from Haemophilus influe
A:Reference number: S27577
A:Accession: S27580
A:Molecule type: DNA
A:Residues: 11-26, 'K', 28-39, 'T', 41-55, 'S', 57-111, 'L', 113-136, 'N', 138-139, 'N', 141-143,
A:Cross-references: EMBL:M94855; NID:9148931; PIDN:AAA24981.1; PID:9148935
A:Experimental source: strain A2

Query Match 7.5%; Score 108.5; DB 2; Length 267;
Best Local Similarity 22.1%; Pred. No. 0.066;
Matches 62; Conservative 51; Mismatches 110; Indels 57; Gaps 13;
QY 1 MONHVISASAAERAAHIAADTFGSG--GIPPOFALMPSELEQAMAEI--VPLISAHP 56
Db 12 LKVIYISLDKDIQRR-----ELFFSQKNTEDVOVSAINTMOKMDLAIAINIEDPKAH- 66
QY 57 YLSGVEK---ACFMSSHAVLMEQALDEG---LPIYAVFEDVLLGEGAEQFLAEEDTLEE 109
Db 67 YFRNATKGEICTLSLSVQKIYEDNDIAEDSYALCEDDALFHSRQONLT--ALISE 124
QY 110 RFDKSAFI-----VLEFMFAKVIYRPDKVLYENRSFPLESEHCIGTAG 155
Db 125 KLEAFITILLGOSKINDENDELEINYPPTFSFLCKRTGDV---NYARP-YKSYPRAGTVG 179
QY 156 YIISREARFELDRPAVLPPEIKAVDLMAFTYFPDKESGMPYQVSPALCTOELHYAKFL 215
Db 180 YLIKSAARFIQIQSQKPFVL-ADDFLLEFQNNINNKYVRLP-----IV 225
QY 216 SONSMLGSDLEKEDREQ-----GRRHRSLLKVMFDLKRAL 249
Db 226 IENPVLLISNLESVRSGLSNNLLKIMKYPFLKKIFAIRKNL 265

RESULT 14

B64623

lipopolysaccharide biosynthesis-associated protein lex2B - Helicobacter pylori (strain 2

C:Species: Helicobacter pylori

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jun-2000

C:Accession: B64623

R:Tombl, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKen-

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weldman, J.M.; Fujii, C.; Bowman, C.; Watney, L.

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: B64623

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-273 <TOM>

A:Cross-references: GB:AE00594; GB:AE00511; NID:g2313957; PIDN:AAD07876.1; PID:g231396

C:Genetics:

A:Start codon: TTG

C:Superfamily: lipopolysaccharide biosynthesis-associated protein

Query Match 7.2%; Score 104; DB 1; Length 273;
Best Local Similarity 23.7%; Pred. No. 0.17;
Matches 64; Conservative 40; Mismatches 80; Indels 86; Gaps 17;

QY 18 IADTFSRGIPF-----QFDALMPSERLQAMAEVPGLSAHRY---- 57

Db 12 VCDTF--GLVREDTTLNLSINATHQAIDAIY-SKTFEGGLHPLVK-KHLHPYFTT 66

QY 58 -----LSGVEK-----ACFMSHAVLMEQALDEGLPYIAVE 88

Db 67 QNKKMGITTNLISESKFYALKYAKFMISGELGCVASHYSLWKCIEIN-FAICILE 125

QY 89 DVLGEGAQFLAEDTWLEERFDKDSAFIVLETMF-AKVIYRDPKVLNENRSEFP--- 144

Db 126 DDTLEKEDKEGFL--DFLEKHI-QELGYIRLMHLLYDASVSEF--LSHKNHEIQERY 178

QY 145 ----LLESEHCAGTAYIISREARFPL--DRFAVLPPERI-----KAVDLMEFTYFF 190

Db 179 GIIRKAYSEGVGTQGYITPKIAKVFLEKCRKVVYPVDTIMDAFTIHGVKNLVLPFVIAD 238

QY 191 DKRGMPVYO----VSPALC-TOELHYAKFL 215

Db 239 DEQISTIAKKEEYSPKIALMRELHF-KYL 267

RESULT 15

E71890

probable lipopolysaccharide biosynthesis protein - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 20-Sep-1999

C:Accession: E71890

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

Ives, C.; Gibson, R.; Werberg, D.; Miller, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: E71890

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-273 <ARN>

A:Cross-references: GB:AE001507; GB:AE001439; NID:g4155326; PIDN:AAD06344.1; PID:g415533

A:Experimental source: strain J99

C:Genetics:

A:Gene: jhp0765

C:Superfamily: lipopolysaccharide biosynthesis-associated protein

Query Match 6.9%; Score 99; DB 2; Length 273;
Best Local Similarity 24.2%; Pred. No. 0.45;

Matches 58; Conservative 36; Mismatches 82; Indels 64; Gaps 14;

QY 30 OFPDALMPSERLQAMAEVPGLSAHRY-----LSGVEK----- 63

Db 38 QIFDAIY-SKTFEGGLHPLVK-KHLHPYITQNIKDMGIATSLISEVSKFYALKYHAKF 95

QY 64 -----ACFMSHAVLMEQALDEGLPYIAVEEDVLLGGAEOFLAEDTWLEERFDKDSAF 117

Db 96 NSLGLGCVASHYSLWKCIEIN-FAICILEDITLKEDEKREGFL--DFLEKHIOE--LG 149

QY 118 IYRLETFMFAKVIYRDPKVLNENRSEFP-----LLESEHCAGTAYIISREARFPL--DR 169

Db 150 YARLMHLLYDASVSEF-LNHNQEIQERYGILKAYSHGVGTQGYITPKIAKVFLEKHSR 208

QY 170 FAVLPPERI-----KAVDLMEFTYFFDKRGMPVYO----VSPALC-TOELHYAKFL 215

Db 209 KVVVPVDTIMDAFTIHGVKNLVLPFVIADDEQISTIAKKEEYSPKIALMRELHF-KYL 267

Search completed: December 2, 2002, 12:00:12
Job time: 14.4266 secs

QY	161	VDLMEFTEPPEKEGPNVQVSPALCTOGLHAKPLFSONSMGSDLEKXREGRHRRSLK	240
DB	161	VDLMEFTFPDEGEGPNVQVSPALCTOGLHAKPLFSONSMGSDLEKDEGRHRRSLK	240
QY	241	VMEFDLRAKLGKFGREKKRMRERORAELEKYGRRVILFK	280
DB	241	VMEFDLRAKLGKFGREKKRMRERORAELEKYGRRVILFK	280
RESULT 2			
ID	LGTE_NEIMB	STANDARD;	PRT; 280 AA.
DT	01-NOV-1997 (Rel. 35, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Lacto-N-neotetraose biosynthesis glycosyl transferase 1gte.		
OS	LGTE OR NMB1926.		
OS	Neisseria meningitidis (serogroup B).		
CC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.		
OX	NCBI_TaxID=491;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MC58 / Serogroup B;		
RX	MEDLINE=96414473; PubMed=8817494;		
RT	Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;		
RT	"Molecular analysis of a locus for the biosynthesis and phase-variable		
RT	expression of the Lacto-N-neotetraose terminal lipopolysaccharide		
RT	structure in <i>Neisseria meningitidis</i> ";		
RL	Mol. Microbiol. 18:729-740(1995).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MC58 / Serogroup B;		
RX	MEDLINE=20175755; PubMed=10710307;		
RA	Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,		
RA	Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.U.,		
RA	Nelson W.C., Gwin M.L., Debay R., Peterson J.D., Hickey E.K.,		
RA	Half D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,		
RA	Mason T.C., Cleek A., Parksey D.S., Blair E., Clifton H., Clark E.B.,		
RA	Cotton M.D., Utterback T.R., Khouri H., Qin H., Yamachyan J.,		
RA	Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L.,		
RA	Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;		
RT	"Complete genome sequence of <i>Neisseria meningitidis</i> serogroup B strain		
RT	MC58";		
RT	Science, 287:1809-1815(2000).		
CC	-1- FUNCTION: ADDS THE FIRST GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN		
CC	IN LOS.		
CC	-1- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS		
CC	STRUCTURE.		
CC	-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on ways		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL: U25839; AAC44086.1; -		
DR	EMBL: AE002541; AAF42255.1; -		
DR	TIGR: NMB1926; -		
DR	InterPro: IPR002654; GT_25.		
DR	Pfam: PF01755; Glyco-transf_25; 1.		
KW	Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;		
KW	complete proteome.		
FT	CONFLICT 16 A -> G (IN REF. 1).		
FT	CONFLICT 89 92 MISSING (IN REF. 1).		
FT	CONFLICT 176 178 EWI -> RVD (IN REF. 1).		
SO	SEQUENCE 280 AA; 32790 MW; 826827E942BA5842 CRC64;		
Query Match	89.2%;	Score 1284;	DB 1; Length 280;
Best Local Similarity	90.4%;	Pred. NO. 1.5e-100;	

	Matches	253;	Conservative	9;	Mismatches	18;	Indels	0;	Gaps	0;
OY	1	MONHVIISLASAERA	RRAHTADTFGSGCIPPOFFDALMPSERLERQAAELVPGLSAHPTYSIG	60						
Db	1	MONHVIISLASAERA	RRAHAIAATGVGIGIEPFOFDALMPSEELNRMAAEIVPLGAOKHLISE	60						
OY	61	VKACFMSHAUWMEALDEGLPYIAVFEDDVLGGCAEQFLAEDTWLEBRDKDSAFIVR	120							
Db	61	VKACFMSHAUWMEALDEGLPYIAVFEDDVLGGDAERFLAEDTWLEBRDKDSAFIVR	120							
OY	121	LETMTAKVIYRPEDKLVTLENRSFPLLESEHGCTAAYIISRREAMPFLDRFVLPERRITKA	180							
Db	121	LETMTAKVIYRPDKLVTLENRSFPLLESEHMGTAGYIISRREAMPFLDRPAVLPREWITKA	180							
OY	181	VDLMVFTEFDEKESMPYIQVSPALCTOELHIAKEPLSQNSMGLSDLEKRPBGRRRRSBLK	240							
Db	181	VDMMVFTEFDEKESMPYIQVNPALCTOELHIAKFLSKNSMLGSLDEKDREORRRRSILK	240							
OY	241	VWFDLKRALGFGREKKRMEROAROLEEKYGRVLIIFK	280							
Db	241	VWFDLKRALGFGREKKRMEROARELEKAYGRVVISFK	280							
	RESULT 3									
	LGBT_NEIGO									
ID	LGBT_NEIGO	STANDARD:	PRT:	279 AA.						
AC	050947;									
DT	01-NOV-1997 (Rel. 35, Created)									
DT	01-NOV-1997 (Rel. 35, Last sequence update)									
DT	15-JUN-2002 (Rel. 41, Last annotation update)									
DE	Lacto-N-neotetraose biosynthesis glycosyl transferase 1gltb.									
GN	LGBT.									
OS	Neisseria gonorrhoeae.									
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.									
OX	NCBI_TaxID=485;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RC	STRAIN=F62;									
RX	MEDLINE=95053752; PubMed=7964493;									
RA	Gotschlich E.C.;									
RT	"Genetic locus for the biosynthesis of the variable portion of									
RU	Neisseria gonorrhoeae lipopolysaccharide.";									
CC	J. Exp. Med. 180:2181-2190(1994).									
CC	-1- FUNCTION: ADDS THE SECOND GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN									
CC	IN LOS.									
CC	-1- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS									
CC	STRUCTURE.									
CC	-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.									
CC	-----									
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration									
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -									
CC	the European Bioinformatics Institute. There are no restrictions on its									
CC	use by non-profit institutions as long as their content is in no way									
CC	modified and this statement is not removed. Usage by and for commercial									
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/									
CC	or send an email to license@sib-sib.ch).									
CC	-----									
DR	EMBL: U14554; AAA68010.1; .									
DR	InterPro: IPR002654; GR_25.									
DR	pfam: PF01755; Glyco.transf.25; 1.									
KW	lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.									
SQ	SEQUENCE 279 AA; 31776 MW; C2219D1A6119D622 CRC64;									
	Query Match	72.0%;	Score 1036.5;	DB 1;	Length 279;					
	Best Local Similarity	71.6%;	Pred. No. 8.1e-80;							
	Matches 209;	Conservative 24;	Mismatches 34;	Indels 25;	Gaps 3;					
OY	1	MONHVIISLASAERA	RRAHTADTFGSGCIPPOFFDALMPSERLERQAAELVPGLSAHPTYSIG	60						
Db	1	MONHVIISLASAERA	RRAHAIAATGVGIGIEPFOFDALMPSEELNRMAAEIVPLGAOKHLISE	60						
OY	61	VKACFMSHAUWMEALDEGLPYIAVFEDDVLGGCAEQFLAEDTWLEBRDKDSAFIVR	120							
Db	61	VKACFMSHAUWMEALDEGLPYIAVFEDDVLGGDAERFLAEDTWLEBRDKDSAFIVR								


```

Db 61 VKACFMSHAVIMEQALDECPYIAVFEDDVLGGAEQFLAEDTWLQERPDPSAFVVR 120
Qy 121 LETMFAYIVRPDKVLYENRSPFLSESHCTAGYIISREARFFLDRAVLPERRIKA 180
    ||||| : | : | : ||||| ||||| : ||||| : ||||| : ||||| : ||||| :
Db 121 LETMFHVITSPSGVADYCGRAFPFLSESHCTAGYIISRKMRFFLDRAVLPERRLHP 180
Qy 181 VDLMEFYFFDEKGMPPYVQVSPALCTQELHYAKFLTSQNSMLGSLDEKREGR----- 233
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 181 VDLMEGNDDEGMPVQCLNPALCAQELHYAKFHDQNSALGSLIEHDLRLNRKQWRDS 240
Qy 234 -----RHRSLSKVMFDLKRALGKFGREKKRMERORQALEVYGRVILFK 260
    : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 241 PANTFKHR-----LIRALTKIGRERKRQROR-----EQLIGKIIVFQ 279

RESULT 4
LCGB_NEIMB
ID LCGB_NEIMB STANDARD: PRT: 275 AA.
AC 051116;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lacto-N-neotetraose biosynthesis glycosyl transferase lgLB.
GN LGTB OR NMB1928.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=96414473; PubMed=8817494;
RA Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;
RT "Molecular analysis of a locus for the biosynthesis and phase-variable
  structure in Neisseria meningitidis."
RL Mol. Microbiol. 18:729-740(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=2015755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
  Eisen J.A., Ketchum K.A., Hood D.W., Beden J.F., Dodson R.J.,
  Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
  Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
  Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
  Cotton M.D., Ulterback T.R., Khouri H., Qin H., Vamathevan J.,
  Gill J., Scarlato V., Maignan V., Pizsa M., Grandi G., Sun L.,
  Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
  MC58."
RL Science 287:1809-1815(2000).
CC -1- FUNCTION: ADDS THE SECOND GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
  IN LOS.
CC -1- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
  STRUCTURE.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U25839; AAC44085.1;
DR EMBL: AE002541; AAF42257.1;
DR TIGR: NMB1928;
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf.25; 1.
KW Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;
  Complete proteome.
FT CONFLICT 97 97 A -> E (IN REF. 1).

```

```

SQ SEQUENCE 275 AA; 31578 MW; E871305E2F6CF70F CRC64;
Query Match 70.3%; Score 1012; DB 1; Length 275;
Best Local Similarity 75.1%; Pred. No. 9e-78;
Matches 202; Conservative 20; Mismatches 43; Indels 4; Gaps 1;

Qy 1 MGNHVISLAAERRAHIAIDFGSGICPPQFDALMPSERLERQAAELVPGISAHPLYSG 60
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 1 MGNHVISLAAERRAHIAIDFGSGICPPQFDALMPSERLERQAAELVPGISAHPLYSG 60
Qy 61 VKACFMSHAVIMEQALDECPYIAVFEDDVLGGAEQFLAEDTWLQERPDPSAFVVR 120
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 61 VKACFMSHAVIMEQALDECPYIAVFEDDVLGGAEQFLAEDTWLQERPDPSAFVVR 120
Qy 121 LETMFAYIVRPDKVLYENRSPFLSESHCTAGYIISREARFFLDRAVLPERRIKA 180
    ||||| : | : | : ||||| ||||| : ||||| : ||||| : ||||| : ||||| :
Db 121 LETMFHVITSPSGVADYCGRAFPFLSESHCTAGYIISRKMRFFLDRAVLPERRLHP 180
Qy 181 VDLMEFYFFDEKGMPPYVQVSPALCTQELHYAKFLTSQNSMLGSLDEKREGR 236
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 181 VDLMEFSDPFDEGMPVQCLNPALCAQELHYAKFHDQNSALGSLIEHDLRLNRKQWRDS 240
Qy 237 RSLKVMFDLKRALGKFGREKKRMERORQ 265
    : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 241 PANTFKHRLIRALTKISRERKRQROR 269

RESULT 5
LCGB_NEIMA
ID LCGB_NEIMA STANDARD: PRT: 279 AA.
AC P57033;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lacto-N-neotetraose biosynthesis glycosyl transferase lgLB.
GN LGTB OR NMA0525.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4a;
RX MEDLINE=20222556; PubMed=10761919;
RA Parthill J., Achman M., James K.D., Bentley S.D., Churcher C.,
  Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
  Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holroyd S.,
  Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
  Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
  Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
  meningitidis Z2491."
RL Nature 404:502-506(2000).
CC -1- FUNCTION: ADDS THE SECOND GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
  IN LOS.
CC -1- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
  STRUCTURE.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AL162753; CAB83817.1;
DR EMBL: AL162753; CAB83817.1;
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf.25; 1.
KW Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;
  Complete proteome.
SQ SEQUENCE 279 AA; 31903 MW; 8703B56513A0D347 CRC64;

```

```

Query Match          67.5%; Score 971.5; DB 1; Length 279;
Best Local Similarity 70.4%; Pred. No. 2,3e-74;
Matches 200; Conservative 24; Mismatches 51; Indels 9; Gaps 2;

OY 1 MONHVSLASAERRAHADTGESRGIPFDPALMPSERLQAMAEVPGLSAHPLYSG 60
DB 1 MONHVSLASAERRAHADTGESRGIPFDPALMPSERLQAMAEVPGLSAHPLYSG 60
OY 61 VERACMSNAHVMEQALDGLPYIAVEEDVLLGEGAEQFLADDTMLEEFDSAFIYR 120
DB 61 VERACMSNAHVMEQALDGLPYIAVEEDVLLGEGAEQFLADDTMLEEFDSAFIYR 120
OY 121 LETMFAKVIYRPDKVLYNENRSEFPLESEHCQTAGYIISREAMRFELDRPAVLPPERIKX 180
DB 121 LETMFAKVIYRPDKVLYNENRSEFPLESEHCQTAGYIISREAMRFELDRPAVLPPERIKX 180
OY 181 VDLMEFTYFDDKGMVYQVSPALCTQELHYAKFLSONSMGSDLEKDR---EGRRHR 236
DB 181 VDLMEFTYFDDKGMVYQVSPALCTQELHYAKFLSONSMGSDLEKDR---EGRRHR 236
OY 237 RSLKVMFDLKRALGKFGREKKRMEERQQAELKVGGRVILEK 280
DB 241 PAWTFKHRLIRALTKTISRERKRORR-----EQLIGKITIVPQ 279

RESULT 6
LIZB_HAEIN
ID LIZB_HAEIN STANDARD; PRT; 266 AA.
AC 057394;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE LIPOILIGOSACCHARIDE biosynthesis protein lic2B.
GN LIC2B.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
NCBI_Taxid=727;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=RM 7004 / Serotype B;
RA High N.J., Jennings M.P., Moxon R.E.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN EXTRACELLULAR LIPOILIGOSACCHARIDES (LOS)
CC BIOSYNTHESIS AND VIRULENCE EXPRESSION. INVOLVED IN THE SYNTHESIS
CC OF THE OLIGOSACCHARIDE MOIETY OF THE LOS MOLECULE BY ADDING GALAC
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC CC EMBL: 254182; CAA09892.1; -
CC DR EMBL: U36398; AAA84947.1; -
CC DR InterPro: IPR002654; GT_25.
CC DR Pfam: PF01755; Glyco_transf_25; 1.
CC KM Transferrase; Glycosyltransferase; Virulence.
CC KW SEQUENCE 266 AA; 31131 MW; 0C68AE9BA249BD39 CRC64;

Query Match          34.8%; Score 500.5; DB 1; Length 266;
Best Local Similarity 40.7%; Pred. No. 6,3e-35;
Matches 114; Conservative 43; Mismatches 98; Indels 25; Gaps 7;

OY 3 NHVSIASAERRAHADTGESRGIPFDPALMPSERLQAMAEVPGLSAHPLYSG 62
DB 4 NHVSIASAERRAHADTGESRGIPFDPALMPSERLQAMAEVPGLSAHPLYSG 62
OY 63 KACFMSHVMQWQCFSEDDLPYIYFEDDVLLGEGAEQFLADDTMLEEFDSAFIYR 122

```

```

DB 63 KACFMSHVMQWQCFSEDDLPYIYFEDDVLLGEGAEQFLADDTMLEEFDSAFIYR 122
OY 123 TMAKVIYRPDKVLYNENRSEFPLESEHCQTAGYIISREAMRFELDRPAVLPPERIKAD 182
DB 123 TMAKVIYRPDKVLYNENRSEFPLESEHCQTAGYIISREAMRFELDRPAVLPPERIKAD 182
OY 183 LMFETFFDDKGMVYQVSPALCTQELHYAKFLSONSMGSDLEKDRQ--GRHRRSIK 240
DB 183 LMFETFFDDKGMVYQVSPALCTQELHYAKFLSONSMGSDLEKDRQ--GRHRRSIK 240
OY 241 VMEFLKRALGKFGREKKRMEERQQAELKVGGRVILEK 280
DB 236 TVL-----ISLAGKPKK-----ILRKIV-RKLFISK 260

RESULT 7
LEX1_HAEIN
ID LEX1_HAEIN STANDARD; PRT; 302 AA.
AC 003974; Q48209;
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1995 (rel. 32, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE LIPOILIGOSACCHARIDE biosynthesis protein lex-1.
GN LEX1 OR LIC2A OR HI0550.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
NCBI_Taxid=727;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Feldman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Georgagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=DL42 / Serotype B;
RX MEDLINE=92065807; PubMed=1956289;
RA Cope L.D., Yergey R., Mertso J., Latimer J.L., Hanson M.S.,
RA McCracken G.H. Jr., Hansen E.J.;
RT "Molecular cloning of a gene involved in lipooligosaccharide
RT biosynthesis and virulence expression by Haemophilus influenzae type
RT B."
RL Mol. Microbiol. 5:1113-1124(1991).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=RM 7004 / Serotype B;
RX MEDLINE=95020659; PubMed=7523834;
RA High N.J., Deadman M.E., Moxon R.E.;
RT "The role of a repetitive DNA motif (5'-CAAT-3') in the variable
RT expression of the Haemophilus influenzae lipopolysaccharide epitope
RT alpha Gal(1-4)beta Gal."
RL Mol. Microbiol. 9:1275-1282(1993).
CC -!- FUNCTION: INVOLVED IN EXTRACELLULAR LIPOILIGOSACCHARIDES (LOS)
CC BIOSYNTHESIS AND VIRULENCE EXPRESSION. INVOLVED IN THE SYNTHESIS
CC OF THE OLIGOSACCHARIDE MOIETY OF THE LOS MOLECULE BY ADDING GALAC.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-11 IS THE INITIATOR.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U32736; AAC22208.1; -
DR EMBL: X56903; CAA40221.1; -
DR EMBL: L19441; AAA65534.1; -
DR PIR: S15282; S15282.
DR TIGR: H10550; -

DR InterPro: IPR002654; GT_25.
KM Pfam: PF01755; Glyco_transf_25; 1.
CC transferase; Glycosyltransferase; Virulence; Complete proteome.
FT VARIANT 29 29 K -> N (IN STRAIN DL42 AND RM 7004).
FT VARIANT 63 70 MISSING (IN STRAIN RM 7004).
FT VARIANT 67 70 MISSING (IN STRAIN DL42).
FT VARIANT 106 106 F -> L (IN STRAINS DL42 AND RM 7004).
FT VARIANT 151 151 Q -> R (IN STRAIN DL42 AND RM 7004).
FT VARIANT 256 256 S -> P (IN STRAIN RM 7004).
FT VARIANT 274 274 N -> D (IN STRAIN RM 7004).
SQ SEQUENCE 302 AA; 35490 MW; BFC204F9B3372D2C CRC64;

Query Match 30.5%; Score 438.5; DB 1; Length 302;
Best Local Similarity 33.0%; Pred. No. 1,2e-29;
Matches 102; Conservative 68; Mismatches 98; Indels 41; Gaps 7;

QY 1 MONHVISLSAERRAHIDTFGSRGIPQFPDAL-----MPSERL 41
DB 4 TENVIYSMNAERKRRHITKOFESKSLSFNFNATYQSIINOSINOSINOSINOSI 63
QY 42 EQAMAE--LVPGLSAHPYLSGVEKACFMSHAVLMEQALDEGLPYIAVEDDVLLEGAE 98
DB 64 NOSINOSINILNIEESRLITLTKGKCLISHFLMNKCVNEMNEFKIPEDDYVIGENNE 123
QY 99 QFLAEDTWLEERFDKDSATVLETWFAKY-IVRPDKVLYNENRSPPLLESEHCAGTIT 157
DB 124 VFLNEMWETKTFDFENDIIRLETFLOPKLEKOTKIPFNSRNDILKSTHWGTAGYI 183
QY 158 ISREARFPLDFPAVLPERRIKAVDLMTYFEDKGMVYOVSPALCTOELHYAKFLSQ 217
DB 184 ISOGAAKYIETLKNIPSEIYAVDELINKNLVDVNTYVQLNPICIQEIQAND---S 240
QY 218 NSMLGSDLEKDEGRHRH-----RSKVMFDLKRALGKFGREKKRMRORQALEK 270
DB 241 KSVLTSGLEKEREKSRKIKRKTLLKQRLTRIKENITRAL-----NRKKMKEDORIKEMQ- 294
QY 271 VYGRVILF 279
DB 295 --GKEIVRF 301

RESULT 8

LPSA_PASHA
ID LPSA_PASHA STANDARD: PRT: 263 AA.
AC 005770;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lpsa protein.
GN LPSA.
OS Pasteurella haemolytica.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Mannheimia.
OX NCBI_TaxID=75985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype A1;
RX MEDLINE=95301198; PubMed=7781993;
RA Potter M.D., Lo R.Y.C.;
RT "Cloning and characterization of a gene from Pasteurella haemolytica
RT A1 involved in lipopolysaccharide biosynthesis.";
RL FEMS Microbiol. Lett. 129:75-81(1995).
RN [2]

RP SEQUENCE OF 189-263 FROM N.A.
RC STRAIN=Serotype A1;
RX MEDLINE=91358346; PubMed=1885539;

RA Abdullah K.M., Lo R.Y.C., Mellors A.;
RT "Cloning, nucleotide sequence, and expression of the Pasteurella
RT haemolytica A1 glycoproteinase gene.";
RL J. Bacteriol. 173:5597-5603(1991).

CC -1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF A MOIETY ON THE CORE
CC OF THE LIPOLYSACCHARIDE MOLECULE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSEMBLED AT THE INNER SURFACE
CC OF THE CYTOPLASMIC MEMBRANE (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U15958; AAA80283.1; -
DR InterPro: IPR002654; GT_25.
KM Pfam: PF01755; Glyco_transf_25; 1.
DR Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.
SQ SEQUENCE 263 AA; 30871 MW; B07C39EC1DA603A7 CRC64;

Query Match 23.0%; Score 331.5; DB 1; Length 263;
Best Local Similarity 35.2%; Pred. No. 9e-21;
Matches 100; Conservative 41; Mismatches 118; Indels 25; Gaps 9;

QY 1 MONHVISLSAERRAHIDTFGSRGIPQFPDALMPSERLQAMAEVLPGLSAHPYLSG 60
DB 1 MANNVISLSAERKRRHITAEFGKONIPQFPDALTPDLIRKAKAFNID--ISNTNLF 58
QY 61 VEKACFMSHAVLMEQALDEGLPYIAVEDDVLLEGAEQFLAEDTWLEERFDKDSAFIVR 120
DB 59 GEIACALSHIALMLAKQONLDYICIFEDDIYVGNNAFEL-----LKTNYIPENHIVK 112
QY 121 LETMAKVIYVRDQVLYN-ENRSPFLSEHCAGTGYITSRAMRFFLDRAVLPERRIK 179
DB 113 LETLPDRIRNRKTEKYLNRRLKLNRSRHVGTAGYILTNNGAEFLN--ILKTLNLP 169
QY 180 AVDLMMFTYFFKEGMPYQVSPALCTOELHYAKFLSONMSGSDLEKDEGRGRHRSLL 239
DB 170 IDDLIFDELYLAKIE-YKVLQNSPALCYQDF-----ILNSTNKSSIQDDRALRCNNEKI 224
QY 240 KVMFDLKR--LGRFGRKKRMRORQALEKYVGRVILF 280
DB 225 KNOAKLKRPNYFLTKIQKELYPFQLOKQ---KVF-TYITFK 263

RESULT 9

Y765_HAETN
ID Y765_HAETN STANDARD: PRT: 282 AA.
AC 057125; 005033;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical glycosyl transferase H10765 (EC 2.-.-.-).
GN H10765.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=9530630; PubMed=7542800;
RA Fleschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Giodek A., Kelley J.M.,


```

GN  ASPC OR AQ_1969.
OS  Aquifex aeolicus.
OC  Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC  Aquifex.
OX  NCBI_TaxID=63363;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-VF5;
RX  MEDLINE=98196666; PubMed=9537320;
RA  Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA  Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjay M., Huber R.,
RA  Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT  "The complete genome of the hyperthermophilic bacterium Aquifex
RT  aeolicus";
RL  Nature 392:353-358(1998).
CC  -!- CATALYTIC ACTIVITY: L-aspartate + 2-oxoglutarate -> oxaloacetate +
CC  L-glutamate.
CC  -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO CLASS-1 OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC  AMINOTRANSFERASES.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL, AE000766; AAC07746.1; -.
DR  HSSP; Q56232; 1BK6.
DR  InterPro: IPR001176; ACC_synthase.
DR  InterPro: IPR004839; Aminotransf1/2.
DR  InterPro: IPR004838; Aminotransf1.
DR  Pfam: PF00155; aminotran_1.2; 1.
DR  PRINTS: PR00753; ACCSYNTHASE.
DR  PROSITE: PS00105; AA_TRANSFER_CLASS_1; FALSE_NEG.
DR  Transfaser: Aminotransferase; Pyridoxal phosphate; Complete proteome.
KM  BINDING: 239 239 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT  BINDING: 239 239 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ  SEQUENCE 394 AA; 43777 MW; E570BAFD080C56E1 CRC64;

Query Match 6.9%; Score 99; DB 1; Length 394;
Best Local Similarity 22.8%; Pred. No. 0.45;
Matches 75; Conservative 38; Mismatches 126; Indels 90; Gaps 17;

OY 9 ASAERRAHIAIDPFG-----SRGIP--PQFPA 34
DB 22 AAKELIRAKGVYIGGAGEPDPTDPFKACIRALREGKTKYAPASAGIPELREAIK 81
OY 35 IMPSERLEQAMALEVPGLSAHPLYSGVERAKCFMSHAVLMEQALDEG-----LPYAVE 88
DB 82 LKENKVEKPESEIYVS-----AGAKNVLF-----IFMALDEDEVLALSPYVATVP 130
OY 89 DVLLEGEGAEQFLAEDTWLEERPKDSAFIVRLTFEFAVIYRPPK-VIN-----Y 138
DB 131 EDIRFEGVVP-----VEVPLKKEKGFQSLSEDKVEKTEKTAIVINSPNPTGAVY 182
OY 139 EKRSPFLSEHCGTAG-YIISREARFPL--DRFVLEPERKKAVIDLMFYY--FDKE 193
DB 183 EEEELKRT-AEFCVERGIFITISDECEYEFVYDAKFVSPASEDEKNTTFVNAASKY 241
OY 194 GMPVYVSPALCTOELHYAKFL-SONSMGSLDEKREGRRHRLSKLWFDLKLGRK- 251
DB 242 SMTGWMLIGVACPEE--YAKVIASLNSQSVNVTTPAIG-----ALEALKNKPSKD 291
OY 252 FGRKKRRMERQQA---ELEKVGRRVI 277
DB 292 FVEMRNFAFERRDYAVEELSKIPGMDV 320

RESULT 12
GLND_RHILQ STANDARD; PRT; 933 AA.
ID GLND_RHILQ

```

```

AC 098C27;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE [Protein-PII] uridylyltransferase (EC 2.7.7.59) (PII uridylyl-
DE transferase) (uridylyl removing enzyme) (UTase).
GN GLND OR ML5321.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MAFF30309;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idehara K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
CC -!- FUNCTION: Modifies, by uridylylation or deuridylylation the PII
CC (glnB) regulatory protein.
CC -!- CATALYTIC ACTIVITY: UTP + [protein-PII] = diphosphate + uridylyl-
CC [protein-PII].
CC -!- SIMILARITY: BELONGS TO THE GLND FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL, AP003006; BAB51794.1; -.
DR  InterPro: IPR002912; ACT.
DR  InterPro: IPR002819; HD.
DR  InterPro: IPR003607; ME_Plpase_HDC.
DR  InterPro: IPR002934; NTP_transf.
DR  Pfam: PF01842; ACT; 2.
DR  Pfam: PF01966; HD; 1.
DR  Pfam: PF01909; NTP_transf_2; 1.
DR  SMART; SM00471; HDC; 1.
KM Transfaser: Nucleotidyltransferase; Nitrogen fixation;
KM Complete proteome.
SQ  SEQUENCE 933 AA; 105279 MW; 79BF065DAC796FP8 CRC64;

Query Match 6.6%; Score 95.5; DB 1; Length 933;
Best Local Similarity 22.1%; Pred. No. 2.6;
Matches 51; Conservative 31; Mismatches 72; Indels 77; Gaps 11;

OY 4 HVISLASAERRAHIAIDPFGSGKIFQFPDALMPSERLQAMALEVPGLSAHPLYSGVEK 63
DB 293 HHMFLTGKAEELH-----FD-----IQRLEAERL-GYTHRPGLSAVNR 330
OY 64 ACFMSHAVLMEQALDEGLYIAVFEDDVLLEGAGAOFLAEDTWLEERPKDSAFIVRLFT 123
DB 331 --FMKH-----YFLVAKD--VG-----DLTRIFCALEE 355
OY 124 MEAKVIYRPDKV-LNVENSFPLSEHCGTAGYIISREARFPLDRFVLEPERKKAVID 182
DB 356 EQAKHVPGFNRILFTFQRRKRL-----AGTSDFYVDNHRINADQVERDP-----VN 405
OY 183 LKMFYTFDEKGMVYQVSPALCTOELHYAKFLSQNSMGLSDEKREGRR 233
DB 406 LRLTFWFAKHGHLFHPDALKLITRSL-----GLVNKSLRDEEANKR 447

RESULT 13
SAD1_MOUSE

```

ID	SAID.MOUSE	STANDARD;	PRF;	627 AA.
AC	060710;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	SAM domain and HD domain-containing protein 1 (Interferon-gamma			
DE	inducible protein Mgl1).			
GN	SAMHD1 OR MGL1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Macrophage;			
RA	MEDLINE=95190404; PubMed=7884320;			
CC	Lafuze W.P., Brown D., Castle L., Zwilling B.S.;			
CC	"Cloning and characterization of a novel cDNA that is IFN-gamma-			
CC	induced in mouse peritoneal macrophages and encodes a putative GTP-			
CC	binding protein.";			
CC	J. Leukoc. Biol. 57:477-483(1995).			
CC	[2]			
CC	SEQUENCE FROM N.A.			
CC	TISSUE=Breast;			
CC	Strausberg R.;			
CC	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- INDUCTION: By interferon gamma.			
CC	-1- SIMILARITY: CONTAINS 1 SAM DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 1 HD DOMAIN.			
CC	-1- CAUTION: Ref.1 sequence differs from that shown due to a			
CC	frameshift in position 576.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U15635; AAA66219.1; ALT_FRAME.			
DR	EMBL; BC012721; AAH12721.1; -			
DR	MGI: 1927468; Samhd1.			
DR	InterPro; IPR002819; HD.			
DR	InterPro; IPR003607; ME_Plase_HDC.			
DR	InterPro; IPR001600; SAM.			
DR	Pfam; PF00536; SAM; 1.			
DR	Pfam; PF01966; HD; 1.			
DR	SMART; SM00471; HDC; 1.			
DR	SMART; SM00454; SAM; 1.			
DR	PROSITE; PS50105; SAM_DOMAIN; 1.			
DR	KW Interferon Induction.			
FT	DOMAIN 46 111 SAM.			
FT	DOMAIN 165 320 HD.			
SQ	SEQUENCE 627 AA; 72650 MW; C68BB653C3F4B17C CXC64;			
Query Match	6.2%; Score 89.5; DB 1; Length 627;			
Best Local Similarity	22.4%; Pred. No. 5.1;			
Matches	54; Conservative 34; Mismatches 76; Indels 77; Gaps 13;			
OY	57 YLSGKRCQFMSHAVLWMOALDEGLPIYAVVEDVL-----LGEAGQFLAEDTW 106			
DB	174 YLAG-----CLV-----IILAERKQPELOISERDILCVQVIGLCHDLGHPFSNHF---- 218			
OY	107 LEERFDKSAFTVRLTFMFAKVIYPPDVVLVNNENSPFLLESEHCGTAGTISRAMPF 166			
DB	219 -----DGRIRP-----ARPEKKWKHEGSIEMFE-----HLVNSNELKLV 254			
OY	167 LDRFAVLPERIKAVDLMMFTYFFDK-----GMPVYQVSPAL-----CTOELHAKFLS 216			
DB	255 MKNYGVLVEEDIT-----FIKEDIMGPPIIPVADSLMPYGRATKTSFLYE--IV 302			
OY	217 ONSMLGSDLEK-DREOGRHRRSLKVMFDLKRALGKFSREKKKKRMRORQALELEVYGR 275			

Db	303	SNKRGGIDVDKWDYARBCHHLGIONNDFDKRFT-KFAR----	ICEVEYKXEDKTYIRK	357
QY	276	V 276		
Db	358	V 358		
RESULT 14				
ID	VNCS_BMDNV	STANDARD:	PRT:	885 AA.
AC	P05842:			
DT	01-NOV-1988 (Rel. 09, Created)			
DT	01-NOV-1988 (Rel. 09, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Putative noncapsid protein NS-1 (Nonstructural protein NS1) . (NCVPI) .			
OS	Bombay denseonucleosis virus (BMDNV) (Silkworm denseonucleosis virus) .			
OC	Virusess; ssDNA viruses; Parvoviridae; Densovirinae; Iteravirus .			
OX	NCBI_TaxID=10809;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Isolate INA;			
RA	MEDLINE=87112952; PubMed=3027382;			
RT	Bando H., Kusuda J., Gojohori T., Maruyama T., Kanase S.:			
RT	"Organization and nucleotide sequence of a denseovirus genome imply a			
RT	host-dependent evolution of the parvoviruses.";			
RL	J. Virol. 61:553-560(1987).			
CC	-1- SIMILARITY: BELONGS TO THE PARVOVIRUSES NONCASPID PROTEIN FAMILY .			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ ,			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL, M15123; AAA67698.1; -.			
DR	PIR: C26796; VCPVE2.			
DR	InterPro: IPR001257; Parvo_NSI.			
DR	Pfam: PF01057; Parvo_NSI; 1.			
KW	Nonstructural protein			
SO	SEQUENCE 885 AA; 100958 MW; 98430617C534D255 CRC64;			
Query Match 6.1%; Score 88.5; DB 1; Length 885;				
Best Local Similarity 19.2%; Pred. No. 9.5;				
Matches 61; Conservative 47; Mismatches 112; Indels 97; Gaps 13;				
QY	17	HIADFGSGRIPIPOFDLMPSERLQMAELVPLGSAHPYISGYEKACFMSHAWLWEOA	76	
Db	601	NVADI -KKAIGTEKKMTRPTAPTTAMSEOPNISA5STSRDINNPIY-----	649	
QY	77	IDEGPIYAVEEDVDVLDEGAEOFL-----AEDTWLEERDKDSAFVRLTWEAKVI	129	
Db	650	VDISLPI---FENNPKVYGIIYDIYDINGNTATYCKCKEKKRPNGLIYASTLKGNV	706	
QY	130	--VAPDKVLNENKSPFLSESHCGTAGYIISREAMFFLDFRAVLPPERIK-----	180	
Db	707	TLPLAPNTIM-----PIGLEN-----GYFSNNGIERRRLTTSVPDDATLAKLNGS	756	
QY	181	-----VDLMFTYFPDK---ESMPYVOVSPALCTOELHAKFLSONSMLG-----	222	
Db	757	ASNNLMAFVDMGYVAFGEOKAPOKSPMPKMGVIRNE-----DNSLLNAKWDILI	808	
QY	223	-----SDLEKDREGRRHRRSLSKMF-----DLKRAIKFGREKKK----	258	
Db	809	KTRIRLTGLSTREVAWATDRIPPOYFKFSQYTGFRYPNINDPTLRSLGTFLKPTKRPGM	868	
QY	259	-----RMRORQAEI 268		
Db	869	DSRTALGELQKORKKML 885		

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 : Search time 24.0515 Seconds
(without alignments)
2398.732 Million cell updates/sec

Title: US-10-007-267-6
Perfect score: 1440
Sequence: 1 MGNHVISLASAERRAHAD.....ERQQAELKVKYGRVILFK 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mnc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.yeast:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriophage:*
17: sp.archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1436	99.7	280	2 Q9REX4	Q9REX4 neisseria g
2	1379	95.8	280	2 Q9RGH8	Q9RGH8 neisseria g
3	1345	93.4	280	2 P96947	P96947 neisseria m
4	1344	93.3	280	2 Q93PR7	Q93PR7 neisseria g
5	1287	89.4	280	2 Q9EVD3	Q9EVD3 neisseria s
6	1282	89.0	280	2 Q93EK4	Q93EK4 neisseria m
7	1046.5	72.7	279	2 Q93PR9	Q93PR9 neisseria g
8	1014	70.4	275	2 Q93EK5	Q93EK5 neisseria m
9	1007	69.9	268	16 Q9JW65	Q9JW65 neisseria m
10	1001	69.5	248	2 Q9RGH1	Q9RGH1 neisseria g
11	984	68.3	275	2 Q9EVD4	Q9EVD4 neisseria s
12	510.5	35.5	266	2 Q8RKU0	Q8RKU0 haemophilus
13	445.5	30.9	280	2 Q9L7A4	Q9L7A4 haemophilus
14	430	29.9	282	2 Q9ZFP8	Q9ZFP8 haemophilus
15	351	24.4	208	16 Q9CLR8	Q9CLR8 pasteurella
16	247.5	17.2	265	16 Q98C26	Q98C26 rhizobium l

17	233	16.2	268	16 Q8YIM5	Q8YIM5 bruceella me
18	171.5	11.9	242	2 Q8RJK5	Q8RJK5 vibrio chol
19	144	10.0	1128	5 Q9VM06	Q9VM06 drosophila
20	139	9.7	230	2 Q8RUL2	Q8RUL2 vibrio chol
21	134	9.3	251	2 Q9ALY2	Q9ALY2 campylobact
22	134	9.3	257	2 Q9ZIS2	Q9ZIS2 escherichia
23	131	9.1	247	2 Q48023	Q48023 haemophilus
24	129.5	9.0	560	4 Q9P226	Q9P226 homo sapien
25	128.5	8.9	517	4 Q9UMW5	Q9UMW5 homo sapien
26	125.5	8.7	207	2 Q8RUM1	Q8RUM1 vibrio chol
27	124.5	8.6	261	2 Q937Y0	Q937Y0 edwardsiell
28	121	8.4	738	4 Q60327	Q60327 homo sapien
29	120.5	8.4	332	16 Q9ZLL7	Q9ZLL7 helicobacte
30	112	7.8	254	2 Q9ALZ3	Q9ALZ3 campylobact
31	111	7.7	231	4 Q8WU19	Q8WU19 homo sapien
32	111	7.7	243	4 Q9H6F1	Q9H6F1 homo sapien
33	109	7.6	261	16 Q9CNC3	Q9CNC3 pasteurella
34	104	7.2	273	16 Q25500	Q25500 helicobacte
35	102	7.1	224	2 Q9X605	Q9X605 actinobacti
36	99.5	6.9	287	2 Q06023	Q06023 haemophilus
37	99	6.9	273	16 Q9ZL17	Q9ZL17 helicobacte
38	98.5	6.8	721	16 Q8Y36	Q8Y36 bruceella me
39	97.5	6.8	224	2 Q47960	Q47960 haemophilus
40	97	6.7	273	2 Q9RHG8	Q9RHG8 helicobacte
41	97	6.7	444	16 Q9ZLL6	Q9ZLL6 helicobacte
42	97	6.7	839	10 Q9M0G3	Q9M0G3 arabidopsis
43	97	6.7	1446	10 Q9SVY9	Q9SVY9 arabidopsis
44	96.5	6.7	257	2 Q9ALY2	Q9ALY2 campylobact
45	95.5	6.6	575	16 Q9KAL7	Q9KAL7 bacillus ha

ALIGNMENTS

RESULT 1

Q9REX4 ID Q9REX4 PRELIMINARY: PRT: 280 AA.
AC Q9REX4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE LGTE.
GN LGTE.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RA Balthazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT "Mutations in the 1gt operon influence serum-resistance in
RT gonococci.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF208063; AAF20992.1; -;
DR InterPro: IPR002654; GT 25.
DR Pfam: PF01755; Glyco_transf_25; 1
SQ SEQUENCE 280 AA; 32448 MW; A20B34CD786942E3 CRC64;

Query Match 99.7%; Score 1436; DB 2; Length 280;
Best Local Similarity 99.6%; Pred. No. 1.2e-119;
Matches 279; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGNHVISLASAERRAHADIDTGSRCIPQFDALMPSERLQNAELVPGISARHYSLG 60
DB 1 MGNHVISLASAERRAHADIDTGSRCIPQFDALMPSERLQNAELVPGISARHYSLG 60
QY 61 VEKACMSHAVILMEQALDGLPIYAFVEEDVLLGEGAEQFLAEDFVLEERFKDSAFYR 120
DB 61 VEKACMSHAVILMEQALDGLPIYAFVEEDVLLGEGAEQFLAEDFVLEERFKDSAFYR 120
QY 121 LETMFAKIVRPDKVLYNENRSPFLLESEHCGTAGYIISREAMRFFLDRAVLPPERIK 180
DB 121 LETMFAKIVRPDKVLYNENRSPFLLESEHCGTAGYIISREAMRFFLDRAVLPPERIK 180

```
OY 181 VDLMMFTYFDKMGMPYQVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSLSK 240
      |||||||
Db 181 VDLMMFTYFDKMGMPYQVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSLSK 240
OY 241 VMDLKRALGKFGREKKKRMERQROALEKVGRRVILFK 280
      |||||||
Db 241 VMDLKRALGKFGREKKKRMERQROALEKVGRRVILFK 280

RESULT 2
O9RGM8 PRELIMINARY: PRT; 280 AA.
AC O9RGM8;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE Glycosyltransferase.
GN LGTE.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1291;
RX MEDLINE=20305049; PubMed=10844691;
RA Harvey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
RA Phillips N.J., Apicella M.A., Balke M.S.;
RT "Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein
RT receptor on human sperm.";
RL EMBL: AF121135; AAF14363.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
KW Transferrase.
SQ SEQUENCE 280 AA; 32304 MW; 2EB5CDE7D2164E6 CRC64;

Query Match
Best Local Similarity 95.0%; Score 1379; DB 2; Length 280;
Best Local Similarity 95.0%; Pred. No. 1.4e-114;
Matches 266; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

OY 1 MONHVISLASAERRAHIDTFSRGIPFOFPDAMPSERLERQAMAEVPGLSAHPLYLSG 60
      |||||||
Db 1 MONHVISLASAERRAHIDTFSRGIPFOFPDAMPSERLERQAMAEVPGLSAHPLYLSG 60
OY 61 VERACMSNAVLMEQALDEGLPYAVFEDDVLGEGAEQFLADTWLEERFDKDSAFIVR 120
      |||||||
Db 61 VERACMSNAVLMEQALDEGLPYAVFEDDVLGEGAEQFLADTWLEERFDKDSAFIVR 120
OY 121 LETMFAKVIYRPDKVLYNERSFPLESEHCGTAGYIISREARFPLDRFAVLPAPERIK 180
      |||||||
Db 121 LETMFAKVIYRPDKVLYNERSFPLESEHCGTAGYIISREARFPLDRFAVLPAPERIK 180
OY 181 VDLMMFTYFDKMGMPYQVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSLSK 240
      |||||||
Db 181 VDLMMFTYFDKMGMPYQVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSLSK 240
OY 241 VMDLKRALGKFGREKKKRMERQROALEKVGRRVILFK 280
      |||||||
Db 241 VMDLKRALGKFGREKKKRMERQROALEKVGRRVILFK 280

RESULT 3
P96947 PRELIMINARY: PRT; 280 AA.
AC P96947;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DE Glycosyl transferase.
GN LGTE.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
```

```
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=126E;
RX MEDLINE=20055626; PubMed=10589709;
RA Jennings M.P., Stikhantha Y.N., Moxon E.R., Kramer M., Poolman J.T.,
RA Kuipers B., van der Ley P.;
RT "The genetic basis of the phase variation repertoire of
RT lipopolysaccharide immunotypes in Neisseria meningitidis.";
RL Microbiology 145:3013-3021(1999).
DR EMBL: U65788; AAB48387.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
KW Transferrase.
SQ SEQUENCE 280 AA; 32589 MW; 1ED681E2E08AFAFC CRC64;

Query Match
Best Local Similarity 93.4%; Score 1345; DB 2; Length 280;
Best Local Similarity 93.6%; Pred. No. 1.5e-111;
Matches 262; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

OY 1 MONHVISLASAERRAHIDTFSRGIPFOFPDAMPSERLERQAMAEVPGLSAHPLYLSG 60
      |||||||
Db 1 MONHVISLASAERRAHIDTFSRGIPFOFPDAMPSERLERQAMAEVPGLSAHPLYLSG 60
OY 61 VERACMSNAVLMEQALDEGLPYAVFEDDVLGEGAEQFLADTWLEERFDKDSAFIVR 120
      |||||||
Db 61 VERACMSNAVLMEQALDEGLPYAVFEDDVLGEGAEQFLADTWLEERFDKDSAFIVR 120
OY 121 LETMFAKVIYRPDKVLYNERSFPLESEHCGTAGYIISREARFPLDRFAVLPAPERIK 180
      |||||||
Db 121 LETMFAKVIYRPDKVLYNERSFPLESEHCGTAGYIISREARFPLDRFAVLPAPERIK 180
OY 181 VDLMMFTYFDKMGMPYQVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSLSK 240
      |||||||
Db 181 VDLMMFTYFDKMGMPYQVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSLSK 240
OY 241 VMDLKRALGKFGREKKKRMERQROALEKVGRRVILFK 280
      |||||||
Db 241 VMDLKRALGKFGREKKKRMERQROALEKVGRRVILFK 280

RESULT 4
O93PR7 PRELIMINARY: PRT; 280 AA.
AC O93PR7;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DE Glycosyl transferase LgTe.
GN LGTE.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PID2;
RA Tong Y., Arking D., Reinhold V., Stein D.C.;
RT "Characterization of lipooligosaccharide structures found in Neisseria
RT gonorrhoeae PID2.";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF313394; AAK70340.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
KW Transferrase.
SQ SEQUENCE 280 AA; 32047 MW; 4F8171AC1BACFB45 CRC64;

Query Match
Best Local Similarity 93.3%; Score 1344; DB 2; Length 280;
Best Local Similarity 93.6%; Pred. No. 1.9e-111;
Matches 262; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

OY 1 MONHVISLASAERRAHIDTFSRGIPFOFPDAMPSERLERQAMAEVPGLSAHPLYLSG 60
      |||||||
Db 1 MONHVISLASAERRAHIDTFSRGIPFOFPDAMPSERLERQAMAEVPGLSAHPLYLSG 60
```

```

QY 61 VERACFMSHAVIMEQALDGLPYIAVFEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
   |||||||
DB 61 VERACFMSHAVIMEQALDGLGVPIAIVFEDVLLGEGAEQFLAEDTWLQERFDKDSAFYR 120
   |||||||
QY 121 LETMFAKVIYRPDKVLYNENRSEFPLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
   ||||| 1: 1: 1: |||||||
DB 121 LETMFAKVIYRPSGVADYGGRAFFPLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
   |||||||
QY 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSIK 240
   |||||||
DB 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSIK 240
   |||||||
QY 241 VMEFDLKRALGKFGREKKRMEORQAELKVGGRVYLTK 280
   |||||||
DB 241 VMEFDLKRALGKFGREKKRMEORQAELKVGGRVYLTK 280
   |||||||

```

RESULT 5

```

Q9EVD3 PRELIMINARY: PRT: 280 AA.
ID 09EVD3:
AC 09EVD3:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE Galactosyl transferase.
GN LGT.
OS Neisseria subflava.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=28449;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21142520; PubMed=11208792;
RA Arking D., Tong Y., Stein D.C.;
RT "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae.";
RL EMBL: AF240672; AAG09766.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
KW Transferase.
SQ SEQUENCE 280 AA; 32762 MW; 85CF77E94F215842 CRC64;

```

```

Query Match      89.4%; Score 1287; DB 2; Length 280;
Best Local Similarity 90.4%; Pred. No. 2, 2e-106;
Matches 253; Conservative 10; Mismatches 17; Indels 0; Gaps 0;

```

```

QY 1 MONHVISLASAERRAHIDTFGRSGIPQFPDALMPSERLQAMAEVPGLSAHPLYSG 60
   |||||||
DB 1 MONHVISLASAERRAHIDTFGARCIPOFPDALMPSERLNMAELVPGIAKHLLSE 60
   |||||||
QY 61 VERACFMSHAVIMEQALDGLPYIAVFEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
   |||||||
DB 61 VERACFMSHAVIMEQALDGLPYIAVFEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
   |||||||
QY 121 LETMFAKVIYRPDKVLYNENRSEFPLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
   |||||||
DB 121 LETMFAKVIYRPDKVLYNENRSEFPLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
   |||||||
QY 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSIK 240
   |||||||
DB 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSIK 240
   |||||||
QY 241 VMEFDLKRALGKFGREKKRMEORQAELKVGGRVYLTK 280
   |||||||
DB 241 VMEFDLKRALGKFGREKKRMEORQAELKVGGRVYLTK 280
   |||||||

```

RESULT 6

```

Q93EK4 PRELIMINARY: PRT: 280 AA.
AC 093EK4:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

```

```

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE LGT.
GN LGT.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M978;
RX MEDLINE=21467954; PubMed=11583844;
RA Zhu P., Klutch M.J., Tsai C.-M.;
RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";
RL FEMS Microbiol. Lett. 203:173-177(2001).
DR EMBL: AF355193; AAL12842.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
SQ SEQUENCE 280 AA; 32898 MW; 504256FF5D92P97 CRC64;

```

```

Query Match      89.0%; Score 1282; DB 2; Length 280;
Best Local Similarity 90.0%; Pred. No. 6e-106;
Matches 252; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

```

```

QY 1 MONHVISLASAERRAHIDTFGRSGIPQFPDALMPSERLQAMAEVPGLSAHPLYSG 60
   |||||||
DB 1 MONHVISLASAERRAHIDTFGRHDIPOFPDALMPSERLNMAELVPGIAKHLLSE 60
   |||||||
QY 61 VERACFMSHAVIMEQALDGLPYIAVFEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
   |||||||
DB 61 VERACFMSHAVIMEQALDGLPYIAVFEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
   |||||||
QY 121 LETMFAKVIYRPDKVLYNENRSEFPLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
   |||||||
DB 121 LETMFAKVIYRPDKVLYNENRSEFPLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
   |||||||
QY 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSIK 240
   |||||||
DB 181 VDLMMFTYFDEKGMVYOVSPALCTQELHYAKFLSONSMGSDLEKDRQGRHRSIK 240
   |||||||
QY 241 VMEFDLKRALGKFGREKKRMEORQAELKVGGRVYLTK 280
   |||||||
DB 241 VMEFDLKRALGKFGREKKRMEORQAELKVGGRVYLTK 280
   |||||||

```

RESULT 7

```

Q93PR9 PRELIMINARY: PRT: 279 AA.
ID 093PR9:
AC 093PR9:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE Glycosyl transferase LGTB.
GN LGTB.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PTD2;
RA Tong Y., Arking D., Reinhold V., Stein D.C.;
RT "Characterization of lipooligosaccharide structures found in Neisseria gonorrhoeae PID2.";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF313394; AAK70339.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
KW Transferase.
SQ SEQUENCE 279 AA; 31637 MW; 7297FL3AFA732379 CRC64;

```

```

Query Match      72.7%; Score 1046.5; DB 2; Length 279;
Best Local Similarity 73.3%; Pred. No. 5.1e-85;
Matches 214; Conservative 18; Mismatches 35; Indels 25; Gaps 3;

```

```
QY 1 MONHVISLASAERRAHIAIDTFCGRIPOFFDAPALMPSERLDQAMAEIVGLSAHPYLSG 60
    |||||
DB 1 MONHVISLASAERRAHIAATFCGRIPOFFDAPALMPSERLDQAMAEIVGLSAHPYLSG 60
QY 61 VERACFMSHVAIVMEQALDEGLPIYIAVEDVLLGEGAEQFLADTWMLEERFDKDSAFIVR 120
    |||||
DB 61 VERACFMSHVAIVMEQALDEGLPIYIAVEDVLLGEGAEQFLADTWMLEERFDKDSAFIVR 120
QY 121 LETMFAKVIVRPDKVLYNERSFPLLESEHCAGTAYIISSEARFFLDRAVILPPERIKR 180
    |||||
DB 121 LETMFAKVIVRPDKVLYNERSFPLLESEHCAGTAYIISSEARFFLDRAVILPPERIKR 180
QY 181 VDLMEFTYFDEKGMPIYQVSPALCTQELHYAFQLSONSMGSDLEKREOGR----- 233
    |||||
DB 181 VDLMEFTYFDEKGMPIYQVSPALCTQELHYAFQLSONSMGSDLEKREOGR----- 240
QY 234 -----RHRRLKTMFDLKRALGFKGREKKRMEORQAELKYYGRVILFK 280
    :||
DB 241 PANTFKR-----LIRALTYIGRERERKRR-----CEOTIGKTIIVPQ 279
```

```
RESULT 8
093EK5 PRELIMINARY: PRT; 275 AA.
AC 093EK5;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
DE LGTB.
GN LGTB.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RX NCBI_TaxID=487;
    [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M978;
RX MEDLINE=21467954; PubMed=11583844;
RA Zhu P., Klutch M.J., Tsai C.-M.;
RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
RT Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";
RL FEMS Microbiol. Lett. 203:173-177(2001).
DR EMBL; AF355193; AL112841.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
SQ SEQUENCE 275 AA; 31577 MW; A73037E0625E7B3 CRC64;
```

```
Query Match 70.4%; Score 1014; DB 2; Length 275;
Best Local Similarity 75.5%; Pred. No. 3.9e-82;
Matches 203; Conservative 21; Mismatches 41; Indels 4; Gaps 1;
```

```
QY 1 MONHVISLASAERRAHIAIDTFCGRIPOFFDAPALMPSERLDQAMAEIVGLSAHPYLSG 60
    |||||
DB 1 MONHVISLASAERRAHIAATFCGRIPOFFDAPALMPSERLDQAMAEIVGLSAHPYLSG 60
QY 61 VERACFMSHVAIVMEQALDEGLPIYIAVEDVLLGEGAEQFLADTWMLEERFDKDSAFIVR 120
    |||||
DB 61 VERACFMSHVAIVMEQALDEGLPIYIAVEDVLLGEGAEQFLADTWMLEERFDKDSAFIVR 120
QY 121 LETMFAKVIVRPDKVLYNERSFPLLESEHCAGTAYIISSEARFFLDRAVILPPERIKR 180
    |||||
DB 121 LETMFAKVIVRPDKVLYNERSFPLLESEHCAGTAYIISSEARFFLDRAVILPPERIKR 180
QY 181 VDLMEFTYFDEKGMPIYQVSPALCTQELHYAFQLSONSMGSDLEKREOGR----- 233
    |||||
DB 181 VDLMEFTYFDEKGMPIYQVSPALCTQELHYAFQLSONSMGSDLEKREOGR----- 240
QY 237 RSLKVMFDLKRALGFKGREKKRMEORQAELKYYGRVILFK 265
    :||
DB 241 PANTFKRRLIRALTISRERERKRRREQ 269
```

```
RESULT 9
09JW65
```

```
ID 09JW65 PRELIMINARY: PRT; 268 AA.
AC 09JW65;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
DE Lacto-N-neotetraose biosynthesis glycosyl transferase (lgth).
GN LGTB2 OR NMA0527 OR LGTB.
OS Neisseria meningitidis (serogroup A), and
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RX NCBI_TaxID=65699; 487;
    [1]
RP SEQUENCE FROM N.A.
RC SPECIES=N.meningitidis (serogroup A);
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,
RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506(2000).
    [2]
RP SEQUENCE FROM N.A.
RC SPECIES=N.meningitidis; STRAIN=A1;
RX MEDLINE=21467954; PubMed=11583844;
RA Zhu P., Klutch M.J., Tsai C.-M.;
RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
RT Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";
RL FEMS Microbiol. Lett. 203:173-177(2001).
DR EMBL; AL162753; CAB83819.1; -.
DR EMBL; AF355194; AL112844.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
SQ Complete proteome.
KW SEQUENCE 268 AA; 30444 MW; 8F43967D053DD02 CRC64;
```

```
Query Match 69.9%; Score 1007; DB 16; Length 268;
Best Local Similarity 72.9%; Pred. No. 1.6e-81;
Matches 204; Conservative 24; Mismatches 40; Indels 12; Gaps 2;
```

```
QY 1 MONHVISLASAERRAHIAIDTFCGRIPOFFDAPALMPSERLDQAMAEIVGLSAHPYLSG 60
    |||||
DB 1 MONHVISLASAERRAHIAIDTFCGRIPOFFDAPALMPSERLDQAMAEIVGLSAHPYLSG 60
QY 61 VERACFMSHVAIVMEQALDEGLPIYIAVEDVLLGEGAEQFLADTWMLEERFDKDSAFIVR 120
    |||||
DB 61 VERACFMSHVAIVMEQALDEGLPIYIAVEDVLLGEGAEQFLADTWMLEERFDKDSAFIVR 120
QY 121 LETMFAKVIVRPDKVLYNERSFPLLESEHCAGTAYIISSEARFFLDRAVILPPERIKR 180
    |||||
DB 121 LETMFAKVIVRPDKVLYNERSFPLLESEHCAGTAYIISSEARFFLDRAVILPPERIKR 180
QY 181 VDLMEFTYFDEKGMPIYQVSPALCTQELHYAFQLSONSMGSDLEKREOGR----- 233
    |||||
DB 181 VDLMEFTYFDEKGMPIYQVSPALCTQELHYAFQLSONSMGSDLEKREOGR----- 240
QY 241 VMDLKRALGFKGREKKRMEORQAELKYYGRVILFK 280
    :||
DB 231 VL--LKRALGKIGREIERARERKRKKLEKHLGRHYVPFE 268
    :||
RESULT 10
09RGNI PRELIMINARY: PRT; 248 AA.
AC 09RGNI;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
DE Glycosyltransferase.
```

GN LGTB.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=485;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1291;
 RX MEDLINE=20305049; PubMed=10844691;
 RA Harvey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
 RA Phillips N.J., Apicella M.A., Balke M.S.;
 RT "Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein
 receptor on human sperm";
 RL Mol. Microbiol. 36:1059-1070(2000).
 DR EMBL: AF121135; AAF14360.1; -
 DR InterPro: IPR002654; GT_25.
 DR Pfam: PF01755; Glyco_transf_25; 1.
 KW Transferase.
 SQ SEQUENCE 248 AA; 28036 MW; D9036FB07D8C9397 CRC64;

Query Match 69.5%; Score 1001; DB 2; Length 248;
 Best Local Similarity 82.2%; Pred. No. 4.9e-81;
 Matches 194; Conservative 15; Mismatches 27; Indels 0; Gaps 0;

OY 1 MONHVISLASAERRAHIDTSGRCIPQFPDAMPSERLBOAMALVPGLSAHPYLSG 60
 DB 1 MONHVISLASAERRAHIDTSGRCIPQFPDAMPSERLBOAMALVPGLSAHPYLSG 60
 OY 61 VERACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKSAFIYR 120
 DB 61 VERACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKSAFIYR 120
 OY 121 LETMFAKVIYRPDKVLYNENRSPFLLESEHCAGTAGYIISREAMRFLDRFAVLPPERIK 180
 DB 121 LETMFAKVIYRPDKVLYNENRSPFLLESEHCAGTAGYIISREAMRFLDRFAVLPPERIK 180
 OY 181 VOLMAMTYFEDEKGMFVYOVSPALCTOELHYAKFLSONSLGSDLEKDEQGRHR 236
 DB 181 VOLMAMTYFEDEKGMFVYOVSPALCTOELHYAKFLSONSLGSDLEKDEQGRHR 236
 OY 181 VOLMAMTYFEDEKGMFVYOVSPALCTOELHYAKFLSONSLGSDLEKDEQGRHR 236
 DB 181 VOLMAMTYFEDEKGMFVYOVSPALCTOELHYAKFLSONSLGSDLEKDEQGRHR 236

RESULT 11
 OQ9EVD4 PRELIMINARY; PRT; 275 AA.
 AC OQ9EVD4;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Galactosyl transferase.
 GN LGTB.
 OS Neisseria subflava.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=28449;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21142520; PubMed=11208792;
 RA Arking D., Tong Y., Stein D.C.;
 RT "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae";
 RL J. Bacteriol. 183:934-941(2001).
 DR EMBL: AF240672; AAG09765.1; -
 DR InterPro: IPR002654; GT_25.
 DR Pfam: PF01755; Glyco_transf_25; 1.
 KW Transferase.
 SQ SEQUENCE 275 AA; 31441 MW; D719F3815F64D14C CRC64;

Query Match 68.3%; Score 984; DB 2; Length 275;
 Best Local Similarity 73.6%; Pred. No. 1.8e-79;
 Matches 198; Conservative 21; Mismatches 46; Indels 4; Gaps 2;

OY 1 MONHVISLASAERRAHIDTSGRCIPQFPDAMPSERLBOAMALVPGLSAHPYLSG 60
 DB 1 MONHVISLASAERRAHIDTSGRCIPQFPDAMPSERLBOAMALVPGLSAHPYLSG 60
 OY 61 VERACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKSAFIYR 120

DB 61 VERACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKSAFIYR 120
 OY 121 LETMFAKVIYRPDKVLYNENRSPFLLESEHCAGTAGYIISREAMRFLDRFAVLPPERIK 180
 DB 121 LETMFAKVIYRPDKVLYNENRSPFLLESEHCAGTAGYIISREAMRFLDRFAVLPPERIK 180
 OY 181 VOLMAMTYFEDEKGMFVYOVSPALCTOELHYAKFLSONSLGSDLEKDEQGRHR--RRS 238
 DB 181 VOLMAMTYFEDEKGMFVYOVSPALCTOELHYAKFLSONSLGSDLEKDEQGRHR--RRS 238
 OY 239 LKYMED--LKRALGKFGREKKRMRORQ 265
 DB 241 PANTFERRRLIALTKISRERKRORREQ 269

RESULT 12
 OQ8RKUO PRELIMINARY; PRT; 266 AA.
 AC OQ8RKUO;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Putative galactosyl transferase Lic2B.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6622;
 RA Pettigrew M.M., Foxman B., Marrs C.F., Gilsdorf J.R.;
 RT "Identification of LOS biosynthesis gene lic2B as a putative virulence
 factor in strains of non-typable Haemophilus influenzae that cause
 otitis medi.";
 RT Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
 RL EMBL: AY091470; AAM12037.1; -
 DR Transferase.
 SQ SEQUENCE 266 AA; 31182 MW; B21CFA372200E2DC CRC64;

Query Match 35.5%; Score 510.5; DB 2; Length 266;
 Best Local Similarity 41.4%; Pred. No. 2.1e-37;
 Matches 116; Conservative 43; Mismatches 96; Indels 25; Gaps 7;

OY 3 NHVVISLASAERRAHIDTSGRCIPQFPDAMPSERLBOAMALVPGLSAHPYLSG 62
 DB 4 NHVVISLASAERRAHIDTSGRCIPQFPDAMPSERLBOAMALVPGLSAHPYLSG 62
 OY 63 KACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKSAFIYR 122
 DB 63 KACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKSAFIYR 122
 OY 123 TMEFAKVIYRPDKVLYNENRSPFLLESEHCAGTAGYIISREAMRFLDRFAVLPPERIK 182
 DB 123 TMEFAKVIYRPDKVLYNENRSPFLLESEHCAGTAGYIISREAMRFLDRFAVLPPERIK 182
 OY 183 LKMFYTFEPKCEGMFVYOVSPALCTOELHYAKFLSONSLGSDLEKDEQGRHR--GRRHR 240
 DB 183 LKMFYTFEPKCEGMFVYOVSPALCTOELHYAKFLSONSLGSDLEKDEQGRHR--GRRHR 240
 OY 241 VMEFLKRALGKFGREKKRMRORQ 265
 DB 241 VMEFLKRALGKFGREKKRMRORQ 265

RESULT 13
 OQ9L7A4 PRELIMINARY; PRT; 280 AA.
 AC OQ9L7A4;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Galactosyltransferase II.

THIS PAGE BLANK (USPTO)